TRANSCRIPTIONAL AND EPIGENETIC REGULATION OF PANCREATIC β-CELL FUNCTION AND DYSFUNCTION

CHAN SU-LING SARAH

BSc. (Honors)

Nanyang Technological University, 2007

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Department of Biological Science
NATIONAL UNIVERSITY OF SINGAPORE
August 2011

ACKNOWLEDGEMENTS

I am extremely blessed to have so many people who have helped me make this journey possible. I would like to thank Dr Lawrence Stanton for giving me the opportunity to work in his lab. He believed that I could accomplish what I set out to achieve, his mentorship has encouraged creativity and flexibility and I have had the chance to grow, learn and explore. I would also like to thank my thesis advisory committee Dr Tara Huber and Dr Thomas Lufkin for feedback along the way.

I am deeply grateful to Dr Han Weiping and Dr Natalia Gustavsson for their invaluable advice on biological and technical aspects of working with β-cells. I am also indebted to Dr Choi Jung Kyoon and Gireesh Bogu whose bioinformatics expertise I have credited in the thesis, and Hwang You Yi who optimized the β-cell dysfunction experiments and was such a pleasure to mentor. I truly appreciate Dr Linda Lim and Dr Akshay Bhinge for brutally honest feedback and countless stimulating discussions, amid much laughter. Special thanks go to Dr Andrew Thomson, with whom many of these initial ideas were debated, and Dr Christina Teh, who instilled in me the foundations of rigorous experimental techniques. I also thank everyone in the GIS Stem Cell group, in particular all past and present members of Stem Cell and Developmental Biology Group 3, for being such fun people to work with. I would like to thank Rani Ettikan for always ensuring we received our items promptly, and being a surrogate mother to everyone. I am incredibly fortunate to have so many wonderful friends: to those I've roamed the world and dived the oceans with, thanks for keeping me sane(r)!

Most importantly, I humbly and gratefully thank my parents and sister for being indescribably giving and supportive over the years. They bore me, raised me, taught me, and loved me. To them, especially my mother, I dedicate this thesis.

TABLE OF CONTENTS

Appendices and Supplementary Tables can be found in the CD attached to the back cover of the thesis.

SUMMARY

Type 2 Diabetes arises from the inability of pancreatic β-cells to compensate for insulin resistance induced by environmental factors in genetically predisposed individuals. Goals in diabetic research are to restore the functionality of damaged β-cells or to replace them, and accomplishing this requires better understanding of the essential determinants of the β-cell phenotype. Pdx1, NeuroD1 and MafA are transcription factors critical in maintaining mature β-cell function. Epigenetic changes to the chromosome architecture in β-cells are associated with the loss of glucose sensitivity and insulin production.

My hypotheses are (1) The transcriptional regulation mediated by Pdx1, NeuroD1 and MafA are key determinants of β-cell function; (2) As β-cell dysfunction develops, chromatin remodeling adversely affects the function of important β-cell transcriptional regulators. To test these hypotheses, I have defined the Pdx1, NeuroD1 and MafA transcriptional regulatory network operative in β-cells, and established the epigenetic profile of normal β-cells and dysfunctional β-cells with impaired glucose-stimulated insulin secretion (GSIS).

Pdx1 binds more independently, and NeuroD1 and MafA have very similar binding profiles. Genes regulated by these transcription factors are over-represented in metabolic, developmental and cellular processes. Combinatorial depletion coupled with combinatorial co-occupancy of these transcription factors shows that genes with Pdx1 binding sites have the most effect on genes that are regulated. When two or three transcription factors are bound, they have the most effect on their target genes. Pdx1, NeuroD1 and MafA regulates genes involved in GSIS: They repress *Itgb1bp2* and *Cplx2* to enhance GSIS, and activate *Tspyl1*, *F13a1* and *Septin7* to dampen the GSIS response, emphasizing the exquisite level of regulation required for the appropriate response of β-cells to a glucose stimulus.

In order to investigate differential epigenetic states in normal and dysfunctional β-cells, I profiled marks of active and repressive chromatin in both cellular states. There is a global increase of open chromatin in dysfunctional β-cells, indicated by the increase in numbers of FAIRE and H3K4me3 peaks and a decrease in H3K27me3 peaks in dysfunctional β-cells. Motif analysis of FAIRE peaks in dysfunctional β-cells show a bias for general transcription factors, while motifs in normal β-cells show enrichments for tissue-specific transcription factors. Genes bound by Pdx1, NeuroD1 and MafA are associated with H3K4me3 in normal β-cells. In dysfunctional β-cells however, there is an increase in the association of these transcription factors and their target genes with the repressive H3K27me3 mark. Bivalent domains marked by both H3K4me3 and H3K27me3 in normal β-cells are enriched for developmental processes and cell regulatory processes. Bivalent chromatin states are modified to different extents in dysfunctional β-cells and lead to expression changes in genes involved in pancreatic development/ function (e.g. *MafB*, *Oxr1*, *Camk2b*, *miR-9*, *Sox17*), genes involved in neuronal development/ function (e.g. *Sema3d*, *Celsr1*), Fox transcription factors (e.g. *Foxf2*, *Foxq1*), and the potassium channels *Kcnc1* and *Kcnq1*.

Little is known about the transcriptional and epigenetic network governing β-cell function and dysfunction. My thesis therefore provides original insight into the Pdx1-, NeuroD1- and MafA- regulated pathways, and presents novel evidence for chromatin remodeling in β-cell dysfunction.

TABLE OF FIGURES

TABLE OF TABLES

CHAPTER 1:

INTRODUCTION

1.1. DIABETES

We are in the midst of a worldwide epidemic of Type 2 Diabetes (T2DM), obesity, and metabolic syndrome, each created by complex interactions between ones genes and the environment in which one lives. Despite advances in treatment, diabetes is the leading cause of chronic renal failure, adult blindness, and limb amputation, and a major risk factor for heart disease, stroke, and birth defects[1]. As a result, the cost of diabetes care in the U.S. currently exceeds $200 billion annually. The immense human and economic costs will increase further as the prevalence of diabetes is expected to affect 380 million worldwide by 2025[2].

The World Health Organization's definition of diabetes is for a single raised glucose reading with symptoms, otherwise raised values on two occasions, of either a) fasting plasma glucose concentration ≥ 7.0 mmol/l (126 mg/dl) or b) a glucose tolerance test, two hours after the oral dose, with a plasma glucose concentration ≥ 11.1 mmol/l (200 mg/dl)[3]. Diabetes is a group of metabolic diseases in which a person has high blood sugar, because the body does not produce enough insulin, and the body's cells do not respond to the insulin that is produced. This high blood sugar produces the classical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger).

1.1.1. TYPE 1 DIABETES MELLITUS (T1DM)

Type 1 Diabetes (T1DM), formally known as insulin-dependent diabetes mellitus (IDDM) or juvenile diabetes, affects an estimated 1 million Americans[4]. It is thought to result primarily from the destruction of pancreatic β cells in the islets of Langerhans by autologous cytotoxic T cells, and leads to insulin deficiency and hyperglycemia (high blood glucose levels). Though it may not be its primary cause, auto-immunity is the predominant effector mechanism of T1DM. T1DM is exercabated in genetically susceptible individuals[5], and the effect of the environment on an individual is also an important one. Epidemiological studies

also suggest a triggering role for enteroviruses, while other micro-organisms might have a protective effect. Therefore prevention of T1DM requires detection at the earliest possible point.

1.1.2. TYPE 2 DIABETES (T2DM)

Unlike T1DM, which is caused by the lack of insulin, T2DM arises from the diminished ability to respond to a given concentration of insulin. This is compounded with the abnormal response of pancreatic β-cells to respond to glucose. Numerous studies show that insulin resistance precedes the development of hyperglycemia in subjects that eventually develop T2DM[6]. However, it is increasingly clear that T2DM only develops in insulin-resistant subjects with the onset of β-cell dysfunction[7]. Both insulin resistance and pancreatic β-cell failure are thought to result from the complex interplay of many different pathways under the combined control of environmental and genetic factors (Figure 1)[8].

The role of genetics in T2DM is indicated by the familial clustering of insulin sensitivity and insulin secretion, the higher concordance rate of T2DM in monozygotic vs. dizygotic twins, and the high prevalence of T2DM in some ethnic groups like the Mexican Americans and Pima Indians[9]. Patterns of inheritance suggest that this is a polygenic and heterogeneous disease, involving multiple genes with different combinations of genes playing a role in different subset of individuals. It is still uncertain exactly how many genes and what the relative contribution of each gene is, and because T2DM is closely linked to other metabolic phenotypes and has a progressive pathogenesis (starting with insulin resistance and β-cell dysfunction, leading to clinical hyperglycemia), elements of complexity are inevitably added to the interpretation of results from genome-wide association studies (GWAS).

Lifestyle factors are also important in the development of T2DM. Non-smokers with high levels of physical activity, a healthy diet (high in fiber, with a high polyunsaturated:saturated

fat) with normal body weight and moderate alcohol consumption, had a 89% lower rate of diabetes[10]. Some other factors that exacerbate the development of T2DM include obesity, hypertension, elevated cholesterol (combined hyperlipidemia), and metabolic syndrome[3] (the group of risk factors that occur together and increase the risk for coronary artery disease, stroke, and T2DM). Other causes include acromegaly, Cushing's syndrome, chronic pancreatitis, cancer, and drugs. Obesity has been found to contribute to approximately 55% of cases of T2DM[11] and is the most potent risk factor for T2DM, possibly by increasing the risk of developing insulin resistance, and decreasing consumption of saturated fats and trans-fatty acids while replacing them with unsaturated fats may decrease the risk[12]. Adipose tissue in the obese releases more of free fatty acids (FFAs), glycerol, hormones, pro-inflammatory cytokines and other factors, that strongly influence insulin resistance and T2DM. Non-essential fatty acids induce insulin resistance and impair ß-cell function, which further impair the capacity to compensate for decreased insulin sensitivity. The normal pancreatic β-cell response to chronic fuel surfeit and obesity-associated insulin resistance is compensatory insulin hyper-secretion to maintain normoglycemia. T2DM only develops in subjects that are unable to sustain the β-cell compensatory response.

While adipose tissue was considered to be the body's fuel depot storing energy in the form of triglycerides, growing evidence suggests that the endocrine function of these tissues in fuel homeostasis can no longer be neglected. The discovery of the adipocyte-secreted peptide hormone leptin led to a paradigm shift favoring this notion[13]. Adipose tissue is known to secrete other molecules like Tumor necrosis factor (TNF)-α, Interleukin (IL)-6, Transforming growth factor (TGF)-β, angiotensinogen, adiponectin, with some of these having been implicated as potential regulators of glucose homeostasis[14]. However, their exact role in contributing to T2DM remains to be explored. Environmental toxins may also contribute to recent increases in the development of T2DM; a weak positive correlation has been found between the concentration of urine of bisphenol A (a constituent of some plastics) and the incidence of T2DM[15].

Given the considerable time and resources it takes to translate basic biomedical discoveries into clinical tools, any current assessment of the clinical value of recent advances in the genetic and biological basis of T2DM is an underestimate. Nevertheless, an improved understanding of fundamental disease mechanisms is already emerging; this will aid in directing future therapeutic advances.

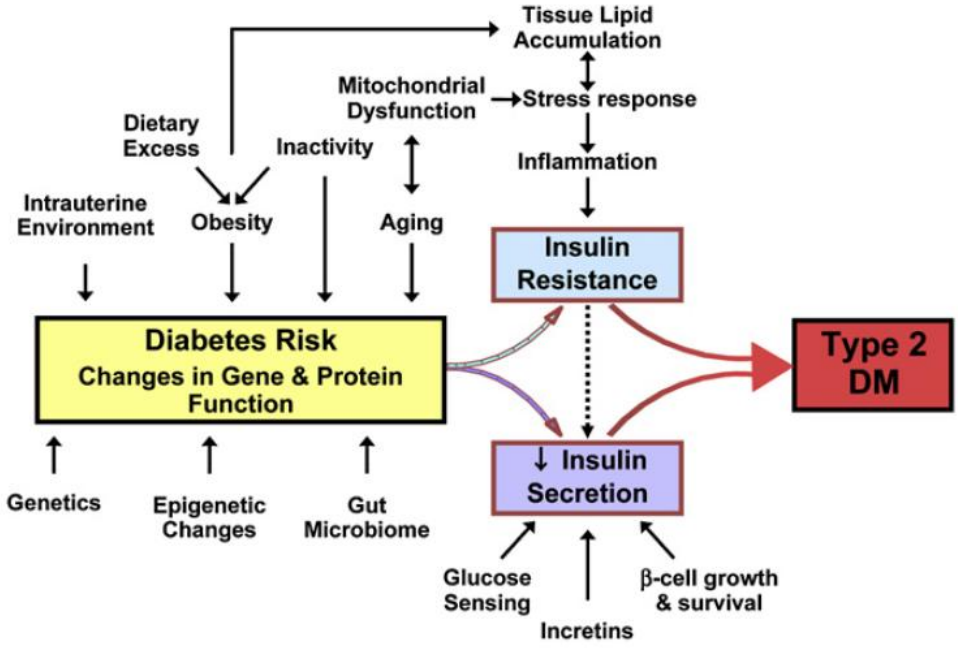


Figure 1. Genetic and environmental factors may influence the risk of diabetes through the pathways illustrated in the figure or through as-yet-unidentified mechanisms affecting insulin sensitivity and/or secretion.

1.2. THERAPY AND TREATMENT IN DIABETES

1.2.1. TREATMENT IN DIABETES

The discovery of insulin in the 1920's was a breakthrough in T1DM treatment, freeing patients from starvation diets that were the mainstream methods of therapy then, prolonging survival for a few weeks, months or in rare cases, a few years[4]. Insulin therapy thus changed diabetes from a rapidly fatal disease to a chronic one associated with significant secondary complications like renal failure, retinopathy, cardiovascular diseases and neuropathy. While aggressive insulin therapy that seeks to maintain glucose homeostasis is the golden aim of diabetes therapy, patients often find it difficult to achieve such fine control and suffer an increased risk of hypoglycaemia[16]. To date, serum biomarkers such as auto-antibodies have been widely used to detect T1DM, and improvements in technologies that rely on metabolomic studies and development of T-cell readouts might enhance diagnostic capabilities. Current clinical trials focus on identifying environmental triggers while therapeutic trials test the efficacy of antigen-specific and antigen-nonspecific immune intervention therapies.

Onset of T2DM can be delayed or prevented through proper nutrition and regular exercise. Intensive lifestyle intervention measures may reduce the risk by over half[17]. Evidence for the benefit of dietary changes alone, however, is limited. In patients with impaired glucose tolerance, diet and exercise and/or metformin (an anti-diabetic drug that reduces hyperglycemia by suppressing glucose production in the liver and increasing sensitivity to insulin)[18] or acarbose may decrease the risk of developing diabetes[17]. Lifestyle interventions are more effective than metformin. Insulin injections may be added to oral medication or used alone. Other classes of medications commonly used to treat T2DM are sulfonylureas and non-sulfonylurea secretagogues (triggers insulin release by inhibiting the K_{ATP} channel in β-cells), alpha-glucosidase inhibitors (to prevent the digestion of

carbohydrates), and thiazolidinediones (to decrease triglycerides and increase high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C))[17].

The way T2DM is managed may change with age. Insulin production decreases because of age-related impairment of pancreatic β-cells and insulin resistance increases because of the loss of lean tissue and the accumulation of fat (especially intra-abdominal fat) and the decreased tissue sensitivity to insulin. Glucose tolerance progressively declines with age, and age-related glucose intolerance is often accompanied by insulin resistance, but circulating insulin levels are similar to those of younger people[19]. Treatment goals for older patients with diabetes vary with the individual, and take into account health status, as well as life expectancy, level of dependence, and willingness to adhere to a treatment regimen.

1.2.2. TRANSPLANTATION THERAPY

While whole pancreas transplantation has demonstrated promising results in the treatment of T1DM, with patients having a >85% likelihood that they will enjoy insulin-independent euglycemia a year later, and about 50% maintaining that favorable metabolic outcome 5 years post-transplantation[4], clinicians are also working to develop isolated islet transplantation techniques as this does not require major surgery. The Edmonton protocol[20] developed in 2000 involves a steroid-free immune-suppressive regimen and high-quality islet cells derived from 2 or more recently deceased donors, and this resulted in a significantly improved graft function. However, an international multi-center trial tested whether the Edmonton protocol could be generalized, and it was found that only 14% of subjects remained insulin-independent 2 years after receiving an islet transplant, and there was a heightened risk of infection, β-cell toxicity and nephropathy[21]. While islet transplantation has the potential to treat diabetic patients, the rate-limiting factor is the availability of islets from organ donors[22]. Therefore, there remains a clear need for an alternative source of cells capable of physiologically regulating insulin secretion.

1.2.3. Alternative Sources of Beta Islets

Transplantation of mature, physiologically-responsive β-cells generated from stem cells is a potential alternative source of islets, and so are xenogenic islets from pigs. The question of whether adult β progenitor cells exist has been a controversial issue for decades, and genetic lineage tracing provided evidence that pre-existing β-cells, rather than stem/progenitor cells, were the major source of new β-cells in adult mice[23]. However, later experiments showed that there exists in the ductal lining, a population of β-cell progenitors that can be activated in the injured adult mouse, and differentiation in situ and in explants of these progenitors give rise to all islet cell types, including β-cells[24]. More recently, rare pancreas-derived multi-potent precursor cells were isolated from adult human islet tissue that were capable of extensive proliferation, self-renewal, and they were capable of generating multiple differentiated pancreatic and neural cell types. These cells also ameliorated diabetes in transplanted mice, demonstrating that the adult mammalian pancreas contains a population of insulin+ multi-potent stem cells[25].

Another alternative source of β-cells would be directed differentiation of embryonic stem cells (ESCs) towards the β-cell lineage. Many attempts have been made to achieve this with various degrees of success, with most studies being only able to generate highly heterogeneous populations of immature β-cells that are not able to respond to environmental stimuli[26]. Recent advances in the field have led to the derivation of pancreatic endoderm from human ESCs[27]. Upon glucose stimulation, these differentiated cells were able to secrete human insulin and c-peptide at levels similar to those of mice that had undergone islet transplantations. They also expressed critical β-cell transcription factors, appropriate processing of proinsulin and the presence of mature endocrine secretory granules, and demonstrated protection against streptozotocin-induced hyperglycemia. The over-expression of three critical pancreatic transcription factors were also able to reprogram adult exocrine cells to β-cells[28]. Porcine islet transplantation is another alternative source of β-cells but it requires major immune-suppressive techniques which may outweigh the risk of diabetes[29].

While all these potential therapies are promising, at this point none of them are widely practiced. It is therefore imperative to understand in greater detail how the mature β-cell phenotype is regulated, and to understand the regulatory mechanisms responsible for glucose-stimulated β-cell insulin secretion, in order to aid our attempts at duplicating its function in vitro and ultimately in patients.

1.3. THE PANCREAS

The adult mammalian pancreas is a heterogeneous organ composed of three major cell types: exocrine acinar cells, ductal cells and endocrine cells[30]. Exocrine cells produce digestive enzymes like trypsiongen, chymotrypsinogen, amylase, lipase that are shunted into the gut via duct cells. The exocrine tissue, organized into dense epithelial acini, makes up >95% of the pancreatic mass. Digestive enzymes are secreted into the acinar lumens which drain into small ducts, and these merge and feed into progressively larger structures, eventually joining the common bile duct. Embedded within the pancreatic exocrine tissue are islets of Langerhans, the endocrine component of the pancreas. The islets of Langerhans are highly-vascularized micro-organs which consists of several endocrine cell types that function together to maintain glucose homeostasis. Histologically, the islets resemble colonies of endocrine tissue suspended within the acinar matrix, as if derived from single progenitor cells; however, lineage tracing shows that each islet is polyclonal in origin[31].

1.3.1. INTER-SPECIES ISLET COMPARISON

Murine islets are the most well studied and are often described as having a highly ordered structure that is 60-80% composed β-cells clustered in a central core, surrounded by 15-20% α-cells, with δ-cells, pancreatic polypeptide (Ppy) cells and ε-cells in the periphery[32]. Human islets tend to contain fewer β-cells and more α-cells compared to rodent islets. The endocrine cells in human islets also do not have a distinct distribution (when compared to rat islets) with α-, β- and δ-cells randomly distributed throughout the islet[33,34], with 50% β-cells, 40% α-cells, 10% δ-cells and few PPY-cells. Small islets appear to be composed of mainly β-cells, whereas large islets have intermingled α- and β-cells[34]. Islet composition varies not only between species, but also within species[34]. Murine islets display a wide array of morphologies under various physiological conditions, including pregnancy, obesity and diabetes. In *db/db* mice, which have a mutation in the leptin receptor and are obese and diabetic, islet architecture and

cellular composition resemble that of humans, suggesting a high degree of plasticity in islet morphology in response to changing physiological conditions[34].

1.3.2. THE UNIQUE BETA CELL FUNCTION

Pancreatic β-cells are able to greatly expand their function and mass in both physiologic and pathologic states of nutrient excess and increased insulin demand, by hypertrophy and proliferation of existing β-cells, increased insulin production and secretion, and formation of new β-cells from progenitor cells[35]. The unique differentiation of the mature β-cell phenotype is responsible for the extraordinary efficiency of these specialized cells in storing and secreting insulin to regulate metabolism in such an exquisite fashion. With successful compensation for insulin resistance, little change in secretory machinery is needed as the distinctive specialization of the β-cell is sufficient for optimal glucose-stimulated insulin secretion (GSIS); the large acute insulin responses to glucose challenge (observed in obese patients) strongly suggest the β-cell phenotype is mostly maintained. The glucose transporter (Glut)-2 allows rapid equilibration between extra- and intra-cellular glucose levels. Glucose is phosphorylated by Glucokinase (Gk) with a K_m of ~8mmol/l, allowing it to function efficiently within the normal range of plasma glucose concentrations[36]. Glucose metabolism occurs primarily by glycolysis, with pyruvate directed to mitochondria for oxidation and little gluconeogenesis[37] or lactate production[38] occurs; all these contribute to maximizing the efficiency of aerobic glycolysis. The specialized mitochondrial shuttles in the β-cell help to contribute to ATP formation (via the glycerol phosphate shuttle that allows reduced nicotinamide dinucleotide (NADH) to be oxidized by the mitochondria, enhancing glycolytic flux)[39] and the generation of NADPH (possibly to enhance insulin secretion) by the malate/ pyruvate shuttle[40].

To maintain this degree of specialization, the genes that are highly expressed in β-cell include those of the secretory products (insulin and islet amyloid polypeptide ((IAPP)), essential

genes for glucose metabolism (*Glut2* and *Gk*), key enzymes for the mitochondrial shuttles (*glycerol phosphate dehydrogenase* (*mGPDH*)) and *pyruvate carboxylase*, and critical transcription factors (Chapter 1.3.6.). At the same time, enzymes that participate in superfluous pathways such as gluconeogenesis and lactate production would be thought to be preferentially suppressed by phosphoenolpyruvate carboxykinase (PEPCK), glucose-6-phosphatase, fructose-1,6-bisphosphatase, lactate dehydrogenase and hexokinase. This is not an exhaustive list of proteins that would have to be expressed and repressed in an active and dynamic manner, in response to the changing environmental milieu.

1.3.3. Synthesis and Processing of Insulin

The liver maintains the baseline level of glucose but β-cells can respond quickly to spikes in blood glucose by releasing stored insulin, while simultaneously producing more. Insulin is initially synthesized as a single-chain 86 amino acid precursor polypeptide, preproinsulin. Subsequent proteolytic processing removes the amino terminal signal peptide, giving rise to proinsulin. Proinsulin is structurally related to insulin-like growth factors (IGF)-I and II, which bind weakly to the insulin receptor. Cleavage of an internal 31-residue fragment from proinsulin generates C-peptide and the A (21 amino acids) and B (30 amino acids) chains of insulin, which are connected by disulfide bonds. β-cells release C-peptide into the bloodstream in equimolar quantities and measuring the levels of C-peptide can give medical practitioners an idea of the viable β-cell mass[41], as it is cleared more slowly than insulin and is a useful marker of insulin secretion, allowing discrimination of endogenous and exogenous sources of insulin in the evaluation of hypoglycemia. C-peptide also helps provide neuropathy and other symptoms of diabetes related to vascular degeneration[42]. β-cells also produce a 37 amino acid peptide called amylin (IAPP) in a 1:100 amylin:insulin ratio. It is the major component of the amyloid fibrils found in the islets of patients with T2DM, and is thought to contribute to glycemic control by inhibiting the appearance of nutrients (especially glucose) into the plasma[43].

1.3.4. GLUCOSE-STIMULATED INSULIN SECRETION (GSIS)

Glucose is the main regulator of insulin secretion by the pancreatic β-cell, although amino acids, ketones, various nutrients, gastrointestinal peptides, and neurotransmitters also influence insulin secretion. Glucose levels > 3.9 mmol/L (70 mg/dL) stimulate insulin synthesis, primarily by enhancing protein translation and processing[44]. Insulin is released in two phases, the first phase release is rapidly triggered in response to increased blood glucose levels, and the second phase release is a sustained, slow release of newly formed vesicles independent of sugar. The first phase of glucose-stimulated insulin secretion begins with its transport into β-cells by Glut2 (Figure 2). The rate-limiting step in GSIS is glucose phosphorylation by Gk to form glucose-6-phosphate, which generates ATP via glycolysis and inhibits the activity of an ATP-sensitive K^+ channel. This channel consists of two separate proteins: one is the binding site for some oral hypoglycemics (e.g. sulfonylureas, meglitinides); the other is Kir6.2, an inwardly rectifying K^+ channel protein. Inhibition of Kir6.2 induces β-cell membrane depolarization, opening voltage-dependent calcium channels that lead to an influx of calcium. This results in activation of phospholipase C, which cleaves the membrane phospholipid phosphatidyl inositol 4,5-bisphosphate into inositol 1,2,4-triphosphate (IP3) and diacylglycerol. IP3 binds to receptor proteins in the endoplasmic reticulum (ER) membrane and allows the release of calcium from ER via IP3-gated channels, further increasing the intracellular concentration of calcium that triggers the exocytosis of insulin-laden vesicles and subsequent insulin secretion.

Insulin exocytosis begins with the synthesis of preproinsulin in the endoplasmic reticulum, followed by budding of immature insulin granules from the trans-Golgi network, insulin granule maturation, and fusion of release-competent dense-core insulin granules with the plasma membrane (Figure 3). A large number of proteins control this secretory pathway but the key molecules controlling membrane fusion are the soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) proteins. Like in other secretory cells, the fusion of the dense-core insulin granules with the plasma membrane involves the assembly of a

complex between the granule membrane Vesicle-associated membrane protein (Vamp)-2 and the integral plasma membrane Syntaxin (Stx)-1a, as well as the membrane-associated Synaptosomal-associated protein (Snap)-25. A large number of accessory factors interact with SNAREs and influence the assembly of the core complex. Insulin secretory profiles reveal a pulsatile pattern of hormone release, with small secretory bursts occurring about every 10 min, super-imposed upon greater amplitude oscillations of about 80–150 min[44].

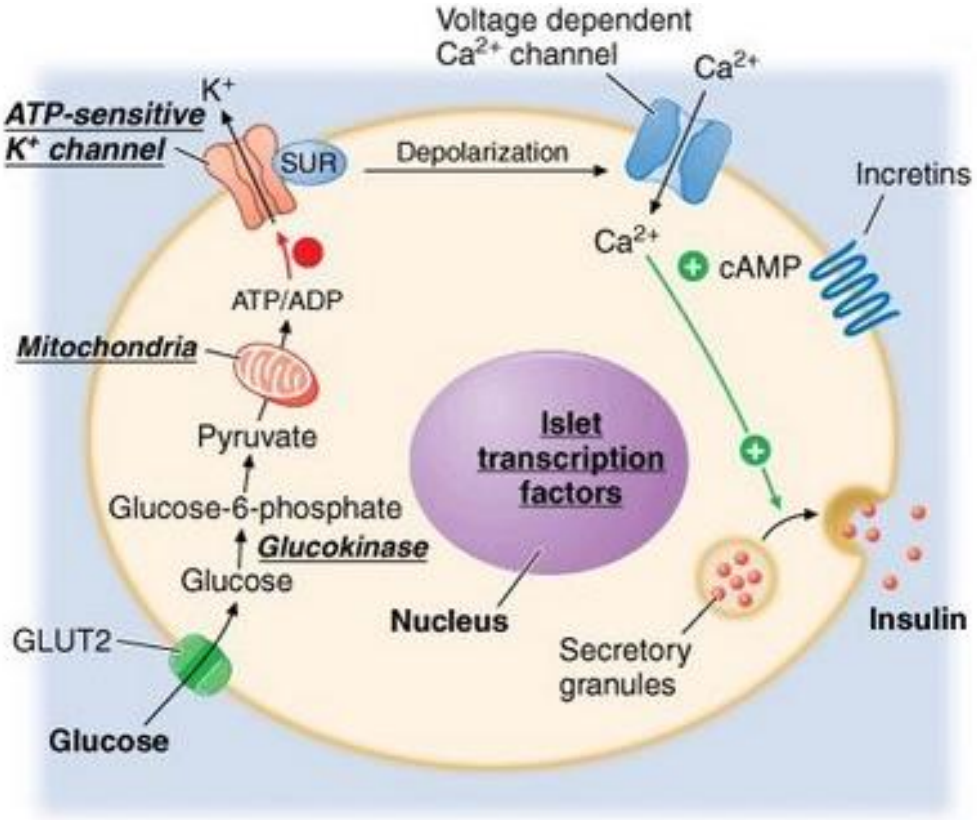


Figure 2. GSIS in the pancreatic β-cell. Glucose is transported into the β-cell by facilitated diffusion through Glut2, which leads to membrane depolarization and an influx of extracellular calcium, resulting in the exocytosis of insulin-containing secretory granules[44].

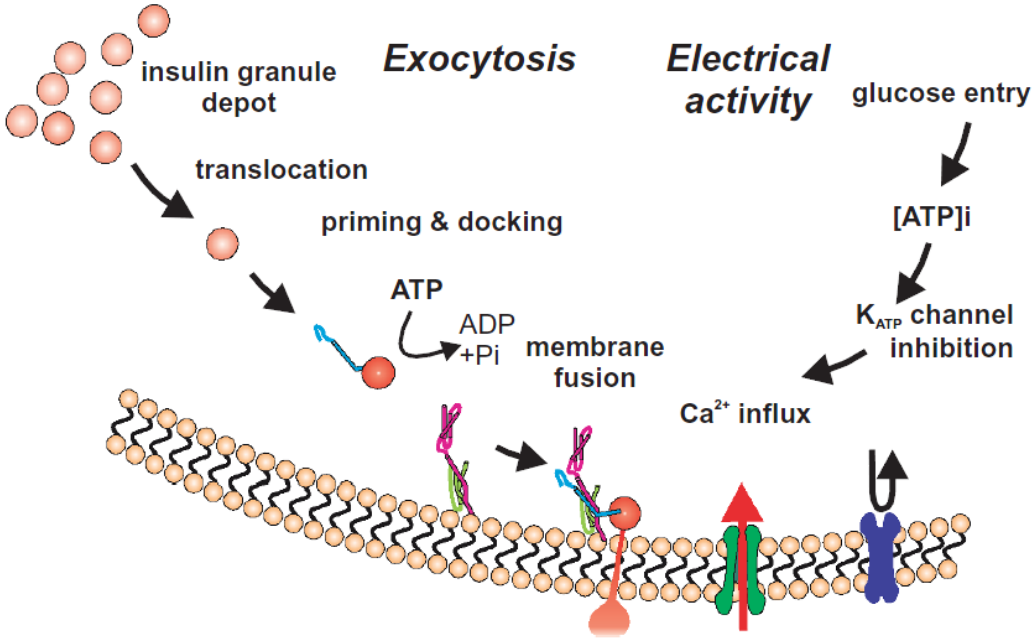


Figure 3. Cellular events involved in GSIS. Reactions to the right (electrical activity) are synonymous with the triggering pathway in β-cell stimulus secretion coupling. Reactions to the left (exocytosis) have the capacity to modify the β-cell secretory response. Some of these reactions are likely to involved in the amplifying action of glucose on insulin secretion[45].

1.3.5. BETA-CELL DYSFUNCTION IN T2DM

Glucotoxicity, lipotoxicity, and glucolipotoxicity are secondary phenomena that are proposed to play a role in T2DM. The underlying concept is that once the primary pathogenesis of diabetes is established (usually involving both genetic and environmental forces), hyperglycemia and hyperlipidemia ensue, exerting additional damaging or toxic effects on the β-cell. In addition to their contribution to the deterioration of function after the onset of the disease, elevations of plasma fatty acid levels that often accompany insulin resistance may, as glucose levels begin to rise beyond the acceptable range, also play a pathogenic role in the early stages of the disease[45]. The word glucolipotoxicity is almost paradoxical, because physiological levels of glucose and lipids are essential for normal β-cell function. Therefore, there exists a spectrum going from normoglycemic and normolipidemic conditions, to that of excess. Glucolipotoxicity states that the alternations in intracellular lipid partitioning underlying the mechanisms of lipotoxicity are dependent upon elevated glucose levels.

In initial stages of insulin resistance due to obesity, β-cell compensation occurs to increase the amount of insulin secreted (Figure 4). Much of this is due to an increase in β-cell mass, probably by β-cell replication and islet neogenesis[35] and enhanced β-cell function[46]. Increased β-cell function has been attributed to increased islet glucose metabolism, FFA signaling (in the form of lipid signaling, lipolysis and triglyceride/ FFA cycling) and sensitivity to incretins like Glucagon-like peptide (Glp)-1. Increased parasympathetic nervous system activity in islets leads to increased insulin secretion[47], and the up-regulation of *Insulin* gene expression (that is linked to the complex interaction between second messenger signals derived from nutrient metabolism, hormonal stimuli, transcription factors and post-translational processing, is implicated in β-cell compensation mechanisms[48].

The loss of GSIS observed in the diabetic state is accompanied by important changes in β-cell phenotype, demonstrated by changes in gene and protein expression in islets from diabetic

Zucker Diabetic Fatty (ZDF) rats[49], and rats following partial pancreatectomy[50]. The alterations in β-cell phenotype has been postulated to be due to dedifferentiation (loss of phenotype), in that the highly expressed genes in β-cells (mentioned above) are down-regulated while those that are normally suppressed are up-regulated in their expression[51]. These cells are also highly stressed, with antioxidant, apoptotic, and pro-apoptotic genes being activated[52], with a concommitant increase in c-myc expression and nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) activation. These changes are consistent with the activation of the pro-inflammatory molecule IL1 in conditions of high glucose[53]. There is also β-cell hypertrophy, which may be a growth process driven by glucose that stops short of cell replication[50]. The loss of the β-cell's ability to respond to glucose might not only be due to an altered phenotype, there could also be important changes in signal transduction pathways that might be unrelated to changes in protein levels.

Insufficient insulin secretion from de-compensated β-cells causes increasingly higher blood glucose levels that continually surround the pancreatic islets, leading to glucose desensitization, β-cell exhaustion and eventually glucose toxicity. Desensitization is a cellular response that protects β-cells from excessive stimulation; repeated challenges by high glucose concentrations rapidly lead to diminished insulin secretory responses to glucose, but not to other agents. On cessation of glucose stimulation, this refractoriness to glucose stimulation eventually recovers and the β-cell once again responds normally to glucose[54]. However, repeated and sustained exposure to high glucose concentrations leads to β-cell exhaustion, as available insulin granules in the exocytotic vesicles have been depleted due to prolonged stimulation of insulin secretion without a sufficient compensatory increase in insulin synthesis. This condition can be "rescued" by allowing the β-cells to rest under conditions of normal glucose levels for a period of time[55]. If the β-cells have been stressed with high levels of glucose for too long a period of time, glucose toxicity occurs[56]. In this state, important mechanisms such as glucose regulation of the *Insulin* promoter and insulin synthesis are unhinged and the β-cell nearly shuts down. Insulin secretion is greatly affected, especially

post-prandially, and the β-cell might even become incapable of producing Insulin mRNA[57] due to the loss of Pancreatic duodenal homeobox (Pdx)-1[58] and v-maf musculoaponeurotic fibrosarcoma oncogene homolog (Maf) A[59].

As nutrient excess persists, hyperglycemia and elevated FFAs negatively impact β-cell function by generating reactive oxygen species, causing alterations in metabolic pathways and increases in intracellular calcium and the activation of endoplasmic reticulum stress. These processes adversely affect β-cells by impairing insulin secretion, decreasing Insulin gene expression and ultimately leading to β-cell death[60].

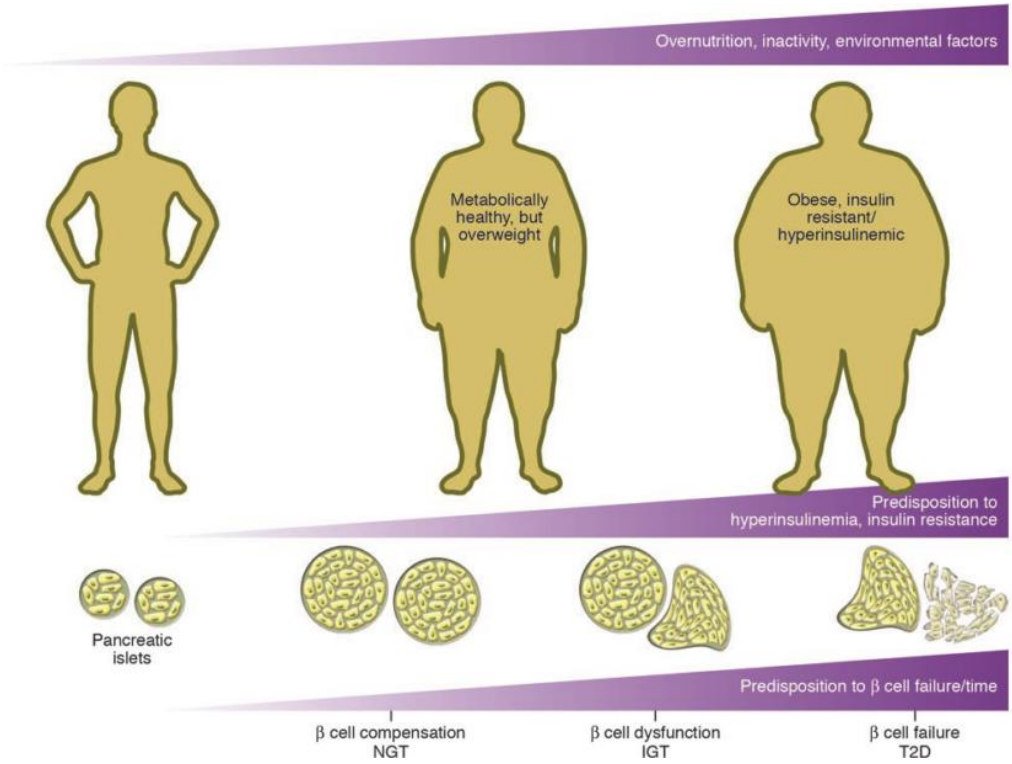


Figure 4. β-cell failure and the natural history of T2D[48]**.** T2D develops in response to over-nutriton and lack of physical activity in subjects that have underlying genetic and acquired predispositions to both insulin resistance (and/or hyperinsulinemia) and β-cell dysfunction. Over time, islet β-cell compensation for the insulin resistance fails, resulting in a progressive decline in β- cell function. As a consequence, subjects progress from normal glucose tolerance (NGT) to IGT and finally to established T2D. Even after diagnosis of T2D, β-cell function continues to worsen such that subjects progress from needing changes in diet/exercise only to requiring oral hypoglycemic agents and eventually insulin for achievement of adequate glycemic control.

1.3.6. *Transcription Factors in Pancreas Development and Function*

Transcription factors orchestrate intricate pathways of cellular growth and differentiation by regulating the rate of transcription of an array of genes in order to maintain cellular homeostasis. Genetic and biochemical studies have begun to unravel the complex cascade of factors that control the proliferation and differentiation of cells in the developing pancreas. The pancreatic β-cell is solely responsible for the transcription of the Insulin gene and the subsequent processing and secretion of insulin in response to increases in extracellular glucose concentrations. The absolute or relative dysfunction of β-cell can have profound metabolic consequences, leading to the development of hyperglycemia and ultimately, diabetes. Therefore, it is imperative to understand the essential determinants of the β-cell phenotype at the physiological and molecular level, and much effort has been put into elucidating the mechanisms that direct the formation of β-cells in the developing embryo and those responsible for the transcription of genes encoding β-cell proteins. Several transcription factors have been identified as being crucial to both of these processes (Table 1). These nuclear regulatory proteins function by binding to characteristic DNA sequences in the upstream regulatory elements of specific genes, and subsequently regulating the rate of transcription of those genes. Transcriptional regulation of genes in the β-cell is manifested by the morphological and phenotypical changes that occur during development, in addition to the synthesis of specific proteins in the mature β-cell, such as Insulin, Glut2, Gk, Neurogenic differentiation (NeuroD)-1, and Iapp.

Transcription factor	Pancreatic cell type	Function
FoxA2	Embryonic endoderm and all islet cell types	Necessary for proper endoderm formation and expression of Pdx1[61]. Also required for regulated insulin secretion in mature β-cells[62].
Hes1	Exocrine progenitors	Directs differentiation of pancreatic precursor cells towards exocrine lineages[63].
HNF1α	β-cells	Activation of insulin and Pdx1[64]. Heterozygous mutations cause MODY in humans[65].
HNF4α	β-cells	Activation of Hnf1 α, insulin and Glut2[66]. Heterozygous mutations cause MODY in humans[67].
Islet1	Pancreatic progenitors, islet-restricted in the adult	Early endocrine cell differentiation.[68]
MafA	Pancreatic progenitors, β-cell-restricted in the adult	Controls insulin gene expression in terminally differentiated β-cells[69].
MafB	Pancreatic progenitors, α-cell-restricted in the adult	Formation of α- and β-cells[70].
NeuroD1	Pancreatic progenitors, islet-restricted in the adult	Islet growth and differentiation. Important for Ins activation in the mature β-cell[71]. Heterozygous mutations cause MODY in humans.
Neurogenin3	Pancreatic progenitors	Directs differentiation of pancreatic precursor cells into endocrine lineages.[72]
Nkx2.2	Pancreatic progenitors, islet-restricted in the adult	Necessary for β-cell precursors to express Nkx6.1[73] and insulin[74].
Nkx6.1	Pancreatic progenitors, β-cell-restricted in the adult	Final differentiation of β-cells[75].
Pax4	Pancreatic progenitors	Formation of β- and δ-cells; represses glucagon transcription. [76]
Pax6	Pancreatic progenitors, islet-restricted in the adult	Formation of α-cells, activates glucagon transcription.[77]
Pdx1	Pancreatic progenitors, β-cell-restricted in the adult	Formation of β- and δ-cells and exocrine tissue[78]. Absence results in lack of pancreas formation[79]. Important activator of Ins[80]. Heterozygous mutations cause MODYin humans[81].

Table 1. Transcription factors important in pancreatic development and function.

1.3.7. TRANSCRIPTION FACTORS IN PANCREATIC DEVELOPMENT

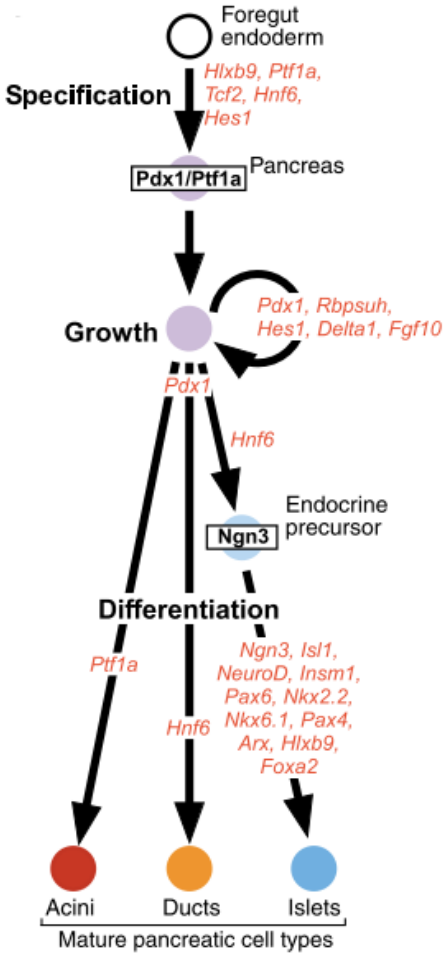
The earliest stages of pancreas organogenesis depend on signaling interactions with surrounding tissues, and the local environment in which the pancreas develops is extremely dynamic regarding the spatial and temporal positioning of the different tissues. Initially, the pancreas primordia are relatively simple structures containing the common pancreatic progenitor cells, but with the appearance of numerous differentiated cell types, gut rotation, and branching morphogenesis, complexity of the organ increases. Before pancreas specification, the formation of definitive endoderm during the gastrulation process is controlled by multiple signaling factors and morphogens where the TGF-β family member nodal seems to be a key molecule in vertebrates[82].

In rodents, the first signs of pancreas organogenesis are the formation, at mid-gestation (embryonic day (E) 9–9.5), of two pancreatic buds, first the dorsal then the ventral, in the region of the posterior foregut endoderm (Figure 5A)[83]. The specification, differentiation and growth of the two pancreatic buds are controlled by various signals originating from the adjacent mesodermal tissues, such as retinoic acid (RA) and the inhibition of sonic hedgehog (Shh) signaling[84]. Dorsally, the pancreatic bud is sequentially exposed in a spatial and temporal fashion to signals from the notochord, dorsal aorta and pancreatic mesenchyme. Ventrally, both the cardiac mesoderm and vitellin veins control pancreas development. The two buds later fuse and the pancreatic epithelium branches and differentiates within the surrounding mesenchyme which further supports pancreatic growth[83]. At the same time, the different pancreatic cell types (α, β, δ, ε and PP) differentiates from immature pancreatic progenitor cells. During their differentiation endocrine precursors delaminate from the epithelium, migrate and aggregate in clusters called islets of Langerhans.

The first indication of pancreas morphogenesis is the up-regulation of the *Pdx1* gene encoding a transcription factor that is essential for development of this organ. *Pdx1*-null mice fail to

develop a pancreas and dies shortly after birth[81]. *Pdx1* mRNA expression is initially detected on embryonic E8.5 in the part of the gut epithelium that later develops into pancreas, and is crucial for the development of endocrine and exocrine cell types. NeuroD1 is detected as early as E9.5 in the pancreatic primordia, indicating that NeuroD1 is present in early precursor cells. Pdx1 is also present in the dorsal and ventral buds as they form around E9.5 and is then expressed throughout the expanding ductal tree up to E13.5, where the basic helix-loop-helix (bHLH) transcription factor Neurogenin (Ngn)-3, determines which of the uncommitted pancreatic progenitor cells are to adopt an endocrine, exocrine or ductal fate (Figure 5B)[85]. *Pdx1* levels decrease at this stage and only reappear in endocrine cells that are destined to become β-cells. At E17.5, *NeuroD1* is mainly expressed in small clusters of endocrine cells. In the adult pancreas, *Pdx1* is expressed mainly in β-cells and is essential in regulating genes involved in maintaining its glucose-responsive phenotype[86]. Other transcription factors like NK homeobox (Nkx)-2.2, Nkx6.1, Paired box (Pax)-4, Pax6, and MafA also contribute to β-cell development[85]. For example, islet progenitor cells can either become precursors of β/δ cells or α-cells, and these alternative fates are promoted by Pax4 and Aristaless-related homeobox (Arx) respectively [87]. Pax6 could act in parallel with Pdx1, Nkx2.2 and Nkx6.1 in some of the β-cells in the late stages of pancreatic development[88]. MafA has been hypothesized to be important for differentiation of β-cells, with Nkx6.1 occuring upstream to MafA during pancreatic development[70].

(A)



(B)

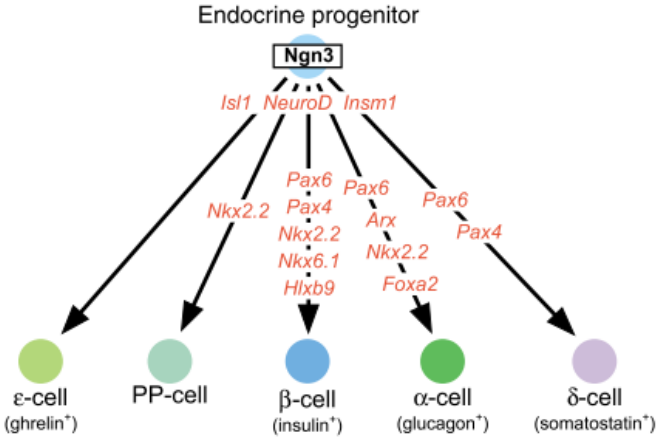


Figure 5. Pancreatic lineages and genes[89]**.** (A) Lineage tracing indicates that all mature pancreatic cell types derive from progenitors that express Pdx1 and/or Ptf1a, and that a subset of these progenitors go on to express Ngn3 and differentiate into islet cells. (B) Pathways of islet-subtype specification. A hypothetical lineage diagram for Ngn3+ endocrine precursors, which give rise to all islet cell types (here, a single Ngn3+ cell is depicted giving rise to each subtype, whereas, in reality, the potential of a given Ngn3+ cell may be more restricted).

1.3.8. TRANSCRIPTION FACTORS IN ADULT PANCREATIC BETA-CELL FUNCTION

In addition to the distinct role of transcription factors in the developing pancreas, many of the aforementioned factors also play a major role in the mature adult β-cell. This is evidenced clinically by the Maturity Onset Diabetes of the Young (MODY) syndromes, in which heterozygous mutations in *Hepatocyte nuclear factor (HNF)-4α*, *HNF1α*, *Pdx1*, *HNF1β* and *NeuroD1* cause progressive impairments in insulin secretion and the eventual development of T2DM[90]. These transcription factors, along with others, often interact together to mediate proper β-cell function that ultimately activates a set of unique β-cell genes, e.g. *Ins*, *Glut2*, *Iapp* or *Gck*).

The usage of Cre–Lox technology in mice has enabled deletions of transcription factors specifically in β-cells, thereby enabling evaluation of their cell-specific roles independently of their roles in development. The β-cell-specific deletion of *Forkhead box (Fox)-a2* leads to the attenuation of the genes encoding *Pdx1* and *ATP-sensitive Kþ channel*[61], thus establishing Foxa2 as an important upstream regulator of these two proteins. Similarly, the β-cell-specific deletion of *Pdx1* in mice results in the attenuation of the genes encoding *Insulin*, *Iapp* and *Glut2*[81]. However, although expression of the target genes in both of these β-cell-specific deletions is clearly diminished, it is not abolished. Therefore, redundancy must exist in the transcriptional network governing gene expression in the β-cell (Figure 6).

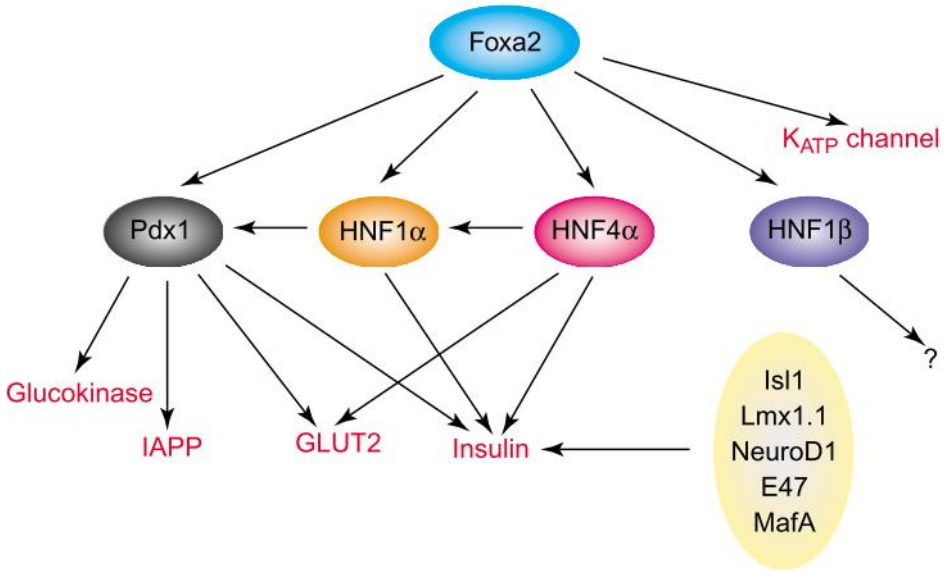


Figure 6. Model for the network of transcription factors that direct β-cell function[85]**.** The major β-cell transcription factors (e.g. Pdx1) are shown as controlling the expression of the genes for other factors or for functionally important proteins within the β-cell (e.g. Glut2). The figure also highlights the redundancy within the network, so that mutation of any one factor (e.g. Hnf4α) serves only to diminish, rather than to eliminate, the expression of the other factors or target genes.

1.3.9. PDX1

Although no single transcription factor can be considered to be the sole determinant of β-cell function, the homeodomain (HD) Pdx1 is the most crucial, with deficits in expression levels resulting in reduced insulin secretion, accelerated β-cell apoptosis, and insulin-deficient diabetes in mice[81]. PDX1 is also clinically important in humans, as different mutations increase susceptibility to the development of hyperglycemia, with a range of phenotypes that include MODY4, T2DM, and gestational diabetes[85]. Pdx1 mediates the adaptive responsiveness of β-cells to changes in extra-cellular glucose concentrations through the regulation of expression levels of glucose sensors, such as the aforementioned Glut2[91] and Gk[92], and through the activation of glucose responsive enhancers within the *Insulin* promoter, in conjunction with other transcriptional regulators[93]. It also controls the expression of Synaptogamin (Syt)-1, Nkx6.1[94], itself[86], and also functions downstream of the insulin signaling pathway in the regulation of β-cell mass[85]. However, the number of target genes is certainly much greater, as *Pdx1*$^{-/-}$ mice display pleiotropic phenotypes[95]. Pdx1 directly regulates *Insulin* transcription through formation of a complex with transcriptional co-activators on the proximal Insulin promoter[96], leading to enhancement of elongation by basal transcriptional machinery. The co-activator p300 interacts with Pdx1 and enhances Insulin transcriptional activity through multiple mechanisms, including the recruitment and activation of components of the basal transcriptional machinery and histone/protein acetylation[80,97,98]. Two highly conserved sequences in the 50-flanking region of the *Pdx1* gene (PH1/area1 and PH2/area2) confer β-cell specific transcriptional activity on a heterologous promoter[99,100]. Pdx1 itself binds to the PH1/area1 element and co-operates with Hnf3β to activate transcription[100]. Pdx1 may be directly activated by the transcription factors NeuroD1[101], Hnf1α and Hnf3β[102]. In T2DM islets, *Pdx1* mRNA expression was increased, despite augmented expression of Foxo1, an inhibitor of *Pdx1* gene expression[103].

1.3.10. NEUROD1

NeuroD1 (a.k.a. β2) is a bHLH transcription factor that plays an important role in pancreas development, and restoration of wild-type gene expression in diabetic mouse models can induce islet neogenesis and reverse the diabetic phenotype[104]. bHLH proteins are transcription factors involved in the determination of cell type and differentiation during development[105], and NeuroD1 is a class B bHLH factor, expressed in pancreatic endocrine cells, the intestine, and brain[106]. Homozygous *NeuroD1*$^{-/-}$ mice are hyperglycemic and exhibit ketonuria, dying within 5 days of birth[71]. These mice have severe diabetes, brought about by the significant reduction in the number of islet cells, because of a failure of endocrine cells to aggregate into mature islets. Mutations in NEUROD1 are linked to MODY6 in humans[1], e.g. a missense mutation of Arg 111 within the DNA-binding domain diminishes the ability of NeuroD1 to bind to DNA and is associated with T2DM in the heterozygous state[1]. Another mutation causes C-terminal truncation of NeuroD1 and leads to a more severe clinical phenotype of T2DM[1]. It functions in a complex with the ubiquitously expressed E47 protein[107] and the heterodimer represents an islet-specific transcription factor that controls both *Insulin*[97] and *Glucagon* (*Gcg*)[108] gene transcription. Over-expressing E47 inhibits *Gcg* gene expression, but activates *Insulin* gene transcription, suggesting that the NeuroD1/E47 ratio may be critical for the regulation of both *Insulin* and *Gcg* expression[108]. NeuroD1/E47 also binds to and activates the upstream *Gk* promoter in the islet[109] and *Islet-specific glucose-6-phophatase catalytic-subunit-related protein*[110]. NeuroD1 may be controlled at the transcriptional and post-trancriptional level by an increase in intracellular Ca^{2+} concentration[111], and the main role of NeuroD1 in the mature β-cell may be to actively repress the *Somatostatin* (*Sst*) promoter in order to maintain *Insulin* expression[112]. NeuroD1 is also implicated in the regulation of insulin secretion via modulation of the expression of ion channel transporters and channels[86] e.g. *Neuronatin* (*Nna*)[113], which is an ion channel transporter, and *Sulfonylurea receptor* (*Sur*)*-1*, which forms potassium channels with Kir6.2[111]. The mis-expression of *NeuroD1* in a human fetal epithelial cell line induced the expression of several genes required for vesicular trafficking and exocytosis, including *Sec24D*, *Snap25*, *Stx1* and *Munc18*[114]. This finding

suggests that NeuroD1 may regulate insulin exocytosis in pancreatic β-cells by directly inducing the expression of genes involved in exocytosis.

1.3.11. MAFA

The C1/RIPE3b element in the *Insulin* gene promoter plays a critical role in β-cell-specific and glucose-regulated expression of *Insulin*[115], and a β-cell-restricted C1-binding factor that appeared in response to glucose in pancreatic β-cell nuclear extracts was originally detected by gel mobility shift analysis and subsequently purified biochemically and identified as MafA[116]. MafA is a basic leucine-leucine zipper (bLZ) transcription factor that is only found in the β-cells of the pancreas and plays an important role in maintenance of β-cell function[117]. Although v-Maf was identified as a retroviral transforming protein[118], it is now widely accepted that Maf family proteins are regulators of a wide variety of cell- and tissue-specific gene expression, and also play a role in cell differentiation during development, e.g. gain- or loss-of-function experiments in the developing chick embryo revealed that L-Maf is a key regulator of lens development[119]. Glucose and oxidative stress regulate *MafA* expression at the transcriptional level through FoxO1, a forkhead transcription factor involved in oxidative stress responses and metabolism[120]. Pdx1, FoxA2 and Nkx2.2 were found to bind to a *cis*-regulatory region of *MafA* approximately 8 kb upstream of the transcription start site, regulating its β-cell-specific expression[121]. In contrast to Pdx1 and NeuroD1, β-cell development does not appear to be affected in *MafA*-deficient mice, although adult mutant animals are glucose intolerant due to diminished *Insulin* transcription and impaired glucose-stimulated insulin secretion, decreased β-cell number and altered islet architecture[43]. The relative unimportance of MafA in β-cell development is surprising, because MafA is produced only in insulin-producing cells of the second and principal phase of β-cell production during mice embryogenesis[42]. The phenotype of *MafA* knockout animals are not as severe as *Pdx1* or *NeuroD1* animals, perhaps because MafA is restricted to β-cells of the pancreas, or because other Maf family members (MafB or C) may compensate for the loss of MafA[32]. MafA appears to also play a role in mediating the expression of *Grauphilin*, *Adiponectin*, *Glut2*,

Nkx6.1, *Pdx1*[122], *pyruvate carboxylase*, *prohormone convertase 1/3*, but it is not known if these genes are *bona fide* MafA targets[32].

1.3.12. Synergistic Activity of Transcription Factors

Pdx1/VP16 expression, together with NeuroD1 or Ngn3 markedly induced *Insulin* gene transcription, ameliorated glucose tolerance, and substantially induced insulin biosynthesis and various β-cell factors necessary for its function, viz., Gk, Sur1, and Kir6.2 in the liver[123]. This triple expression was more effective than when the genes were ectopically over-expressed on their own, possibly due to marked induction of the *Insulin* mRNA expression to increase *Insulin* promoter activity by an auto-regulatory loop[124]. Another possibility is that the transcription factors are recruited to the *Insulin* promoter region by MafA, and together with Pdx-1 and NeuroD1, this enables these transcription factors to exert strong synergistic effects and to markedly induce *Insulin* gene expression[124]. Pdx1, E47 and NeuroD1 interacts synergistically to stimulate *Insulin* promoter activity[107]. Pdx1, MafA, and NeuroD1/E47 acting through the *Insulin* promoter proximal A1, C1, and E1 sites respectively, play an important role in maintaining basal promoter activity of the human INS gene, with evidence of the transcription factors having a synergistic effect on the human INS promoter[125]. Pdx1 had the strongest stimulatory effect on the *Insulin* promoter, relative to the weaker effects of MafA, NeuroD1, and E47. When the human INS promoter was modified to resemble the rat *Insulin1* promoter, there was more synergistic activation generated[100], emphasizing the differences in regulation of the human promoter and the well-characterized rodent *Insulin* promoter. Pdx1, Pax6, NeuroD1, and Nkx 2.2 regulate the *Insulin* gene by binding regulatory control regions of other genes selectively transcribed in islet cells[126]. MafA functionally interacts with Pdx1 and NeuroD1 to promote synergistic activation of the *Insulin* enhancer-driven reporter cell line in non β-cells[127], and MafA, NeuroD1/E47, and Pdx1 cyclically bind to the mouse *Insulin2* promoter[128]. Pdx1 and NeuroD1 co-occupy the promoters of not just the *Insulin* gene, but many other genes involved in secretion, transporters, transcription and microRNAs (miRNAs) involved in maintaining β-cell function[86], suggesting that Pdx1 and NeuroD1 co-

regulate a significant number of genes. Ngn3, Pdx1 and MafA were able to reprogram differentiated pancreatic exocrine cells in adult mice to cells very similar to β-cells, these β-cells were indistinguishable from endogenous islet β-cells in size, shape and ultrastructure, expressing genes essential for β-cell function could ameliorate hyperglycaemia by remodelling local vasculature and secreting insulin[28].

1.4. EPIGENETIC REGULATION IN PANCREATIC BETA-CELLS

Although obesity, reduced physical activity, and aging increase susceptibility to T2DM, many people exposed to these risk factors do not develop the disease. Recent genome-wide association studies have identified a number of genetic variants that explain some of the inter-individual variation in diabetes susceptibility[129]. There is also increasing evidence suggesting a role for epigenetic factors in the intricate interplay between genes and the environment, especially in complex diseases like T2DM[130]. Epigenetics is defined as heritable changes in gene function that occur without a change in the nucleotide sequence, and these modifications of the genome provide a mechanism that allows the stable propagation of gene activity states from one generation of cells to the next[131]. Epigenetic states can be modified by environmental factors, which may contribute to the development of abnormal phenotypes. In plants, it is well established that epigenetic modifications can be inherited from one generation to the next[132]. There are at least two main classes of epigenetic information that can be inherited with chromosomes: one involving changes in chromatin proteins (usually involving modification of histone tails), and the other involving changes in DNA methylation.

In eukryotes, DNA is assembled with histones to form the nucleosome, in which 147 base pairs of DNA is wrapped approximately two turns around an octameric complex composed of two molecules of each of the four histones H2A, H2B, H3, and H4. The core histones are densely packed but their NH2-terminal tails can be modified by acetylation, methylation, glycosylation, sumoylation, phosphorylation, and ADP ribosylation[133]. These modifications are important for determining the accessibility of the DNA to the transcription machinery as well as for DNA replication, recombination, and higher-order chromosomal organization. The most common moditications involve acetylation and methylation of lysine residues in the H3 and H4 amino termini. Histone deacetylases (HDACs) remove and histone acetyl transferases (HATs) add acetyl groups to lysine residues on histone tails[133]. HAT activity and increased histone acetylation usually correlate with increases in gene transcription; an initial explanation for how that happened was that histone acetylation removed the positive charges of the native

lysine residues on histone tails, abolishing the binding of negatively charged DNA, resulting in a relaxed chromatin state that facilitated access to transcriptional machinery, allowing transcription[133]. Histone acetylation may also recruit bromodomain proteins that can act as transcriptional activators[134]. However, recent work has led to the proposal of different models, for example the histone code hypothesis which suggests that multiple histone modifications act in combination to regulate transcription[135]. H2A.Z associates with functional regulatory elements, and the insulator CTCF marks boundaries of histone methylation domains[136].

1.4.1. HISTONE METHYLATION

Histone methylation is associated with both transcriptional repression and activation, depending on the degree of methylation and the specific lysine and/or arginine residue modified, and lysine residues can be mono-, di-, or tri-methylated in vivo, providing an added complexity of transcriptional regulation[134]. The mono-methylations of H3K27, H3K9, H4K20, H3K79, and H2BK5 are all linked to gene activation, while tri-methylations of H3K27, H3K9, and H3K79 are linked to repression[136]. High levels of H3K4me1, H3K4me2, and H3K4me3 are detected surrounding transcription start sites (TSSs), while H3K36me3 peaks are near the 3' end of genes. The apparently opposite modifications H3K4me3 and H3K27me3 have been found to co-localize in ''bivalent domains'' in ESCs[137], and these have been suggested to function in the differentiation of these cells, as these "bivalent domains" have been found in developmental regulators. Recent evidence has confirmed the presence of these domains in mouse embryonic fibroblasts (MEFs), neural progenitors[138] and human T cells[139], proving that these "bivalent domains" are not an ESC-specific phenomenon. A comparison of H3K4me3 and H3K27me3 in islets show that these "bivalent domains" are usually mutually exclusive or absent, but there are few genes that had high levels of both marks, e.g. the *Homeobox* (*Hox*) gene clusters and neuronal transcription factors, perhaps indicating a retention of "memory" of earlier transcription or the re-activation of these genes under certain metabolic conditions[140].

1.4.2. DNA METHYLATION

The second major class of epigenetic regulation is DNA methylation, where a cytosine base is modified by DNA methyltransferase (DNMT) at the C5 position of cytosine. Approximately 70% of CpG dinucleotides in human DNA are constitutively methylated, while most of the unmethylated CpGs are located in CpG islands, which are CG-rich sequences located near coding sequences that act as promoters for the genes that are associated with them. Approximately half of mammalian genes have CpG islands. Methylation of CpG sites is maintained by DNMTs: DNMT1 maintains DNA methylation during replication[141] while DNMT3a and b are involved in *de novo* DNA methylation[142]. DNA methylation is commonly associated with gene silencing and is achieved by modifying the binding affinity of transcription factors that are essential for activating gene transcription, or by recruiting proteins that bind specifically to methylated CGs (methyl-CG-binding proteins e.g. MeCP2). These in turn further recruit HDACs and co-repressors like Sin3A[143]. This silencing effect contributes to X-chromosomal inactivation, genomic imprinting, and transcriptional regulation of tissue-specific genes during development and differentiation[131]. Most CpG islands remain unmethylated in normal cells, but in certain conditions e.g. cancer[144] and oxidative stress, previously unmethylated CpG sites can become *de novo* methylated. This anomalous methylation is accompanied by local changes in histone modification and chromatin organization, resulting in the CpG island and its associated promoter taking on a repressed conformation which is incompatible with gene transcription. It is not known what makes particular CpG islands susceptible to aberrant methylation. A "sequence signature associated with aberrant methylation" is proposed to exist[145], where Pdx1 was identified as being 1 of 15 CpG genes (a total of 1749 genes with CpG islands were examined) that was methylation-susceptible under conditions of increased methylation induced by over-expression of a DNMT.

Although DNA methylation and histone modification are carried out by different chemical reactions and require different sets of enzymes, there seems to be a biological relationship between the two systems that plays a part in modulating gene repression programs in mammalians. It appears that this relationship can work in both directions: Histone methylation can direct DNA methylation patterns, and DNA methylation can simultaneously serve as a template for some histone modifications after DNA replication (Figure 7)[146]. At the molecular level, these connections might be accomplished through direct interactions between histone methyltransferases (HMTs) and DNMTs. For example, methylation of lysine 9 on H3 promotes DNA methylation, while CpG methylation stimulates methylation of lysine 9 on H3[147]. Therefore chromatin modifications induced by adverse stimuli are self-reinforcing and can propagate.

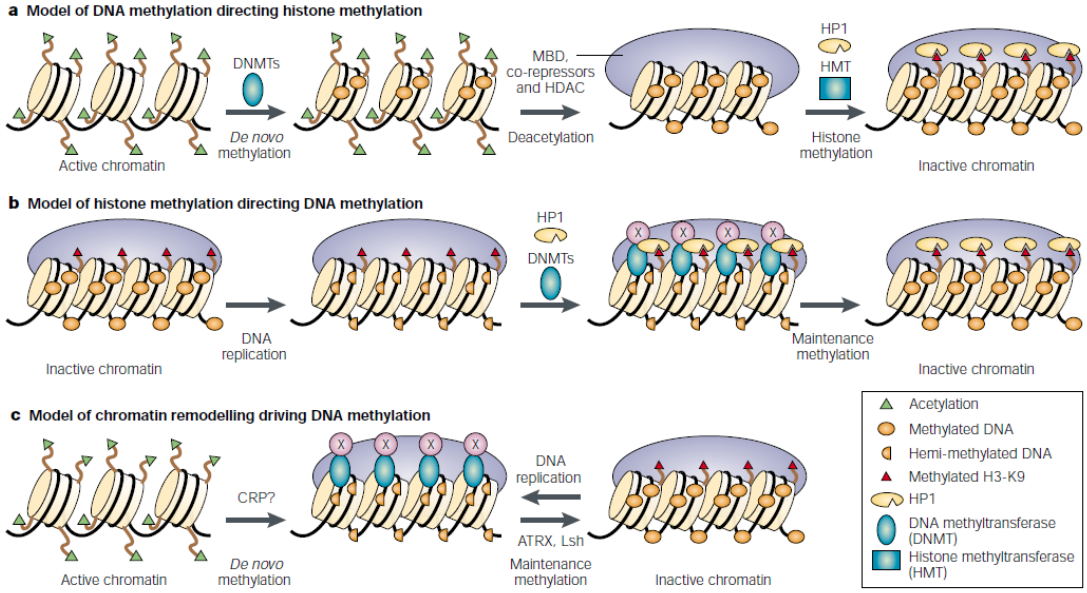


Figure 7. The link between DNA methylation, histone modification and chromatin remodelling[148]**.** In mammalian cells, both DNA methylation and histone modification are involved in chromatin silencing. DNA methylation and histone modification are believed to be interdependent processes. Three possible models of how they might influence each other are shown. (A) A model of DNA methylation directing histone methylation. DNA methylation patterns are established through *de novo* methylation by DMT3A and DNMT3B, and are maintained by DNMT1. Methyl-CpG-binding proteins (MBD) and HDAC complexes, such as the MECP2–Sin3a–HDAC complex, are believed to then be recruited to the methylated region to induce histone deacetylation and silencing. The chromatin then attracts HMTs such as Suv39h or G9a, which methylate the lysine 9 residue on histone H3 (H3K9) and stabilize the inactive state of the chromatin. (B) A model of histone methylation directing DNA methylation. Methyl H3K9 acts as a signal for inactive chromatin by recruiting HP1 to methylated histones, which might in turn recruit DNMTs directly or indirectly (through an unknown factor, factor X) to the silent chromatin to maintain DNA methylation and stabilize the inactive chromatin. (C) A model of chromatin remodelling driving DNA methylation. The ATP-dependent chromatin-remodelling and DNA-helicase activities of proteins, such as ATRX and Lsh, might facilitate DNA methylation and histone modification by unwinding nucleosomal DNA to increase its accessibility to DNMTs, HDACs and HMTs. The disrupted function of these proteins impairs both DNA methylation and histone methylation, as has been shown in plants.

1.4.3. EPIGENETICS, THE ENVIRONMENT & T2DM

Initial observations suggesting a role for environmental cues in establishing an epigenetic mark came from studies of the agouti mouse model, where a retrotransposon element containing a strong promoter is inserted close to the promoter of the gene encoding the Agouti protein. If the retrotransposon promoter sequence is methylated, the offspring are wild-type agouti because it cannot override the agouti promoter, and vice versa. Interestingly, yellow mothers produce more offspring than agouti mothers, but their offspring suffer from obesity, hyperinsulinaemia and diabetes[149]. Maternal diet alters the phenotype of these mice: when pregnant females are fed a methyl-donor-rich diet, a larger proportion of offspring have a wild-type coat color (indicating that methylation has silenced the offspring's retrotransposon, allowing the wild type agouti promoter to be expressed) in comparison to mothers fed a standard diet, showing that maternal nutrition can change the stable expression of genes in offspring through epigenetic modifications occurring *in utero*[150].

In the context of T2DM, pregnant women exposed to famine conditions during their late stage pregnancy in 1944/5 gave birth to smaller babies with an increased risk of developing insulin resistance[151], and these observations have been repeated in rodent models of intrauterine growth retardation (IUGR). The effects of neonatal leptin (a hormone that regulates energy intake and expenditure) on hepatic gene expression and epigenetic status in adulthood were found to be directionally dependent on the animal's nutritional status *in utero*. This showed that during mammalian development, the direction of the response to one cue can be determined by previous exposure to another, suggesting the potential for a discontinuous distribution of environmentally-induced phenotypes (an epi-phenotype)[152].

One of the key environmental perturbations associated with T2DM is exposure to an adverse intrauterine milieu, such as IUGR which induces mitochondrial dysfunction in the β-cell, leading to increased production of reactive oxygen species (ROS) and oxidative stress[153].

Uteroplacental insufficiency (the most common cause of IUGR in the developed world) induces epigenetic modifications in offspring: pups are born spontaneously with decreased levels of glucose, insulin, Igf1 and amino acids[154]. Diabetes develops in these animals at 15–26 weeks of age with underlying β-cell secretory defects and insulin resistance, epigenetic modifications affecting processes important to glucose regulation and insulin secretion have been described in the IUGR liver, pancreatic β-cells and muscle[154].

The 6q21 region / HDAC2 gene is significantly associated with diabetes[155]. Another locus identified in a genome wide scan of type 2 diabetes lies on chromosome 19q13 and the HDAC SIRT2 maps there[156]. In mice heterozygous for the HAT cyclic adenosine monophosphate (cAMP)-response element (Cre)-binding protein (CBP) demonstrate increased insulin sensitivity and glucose tolerance even while demonstrating a marked lipodystrophy of white adipose tissue[157]. Exposure to oxidative stress can directly mediate both DNA methylation and remodelling of chromatin in multiple disease models, perhaps explaining how aberrant epigenetic programming leads to T2DM[158]. Random DNA methylation changes that occur during aging in several tissue types are also associated with increased oxidative stress, affecting the expression of multiple genes[158]. Because of their abundant lysine residues, histones are susceptible to oxidative stress[159]. The discovery of histone demethylases (e.g. the Jumonji C family) link epigenetic processes to oxygen gradients during development[160]. This is particularly relevant to T2DM as there are now substantial data showing that oxidative stress plays a significant role in β-cell deterioration.

Therefore, the molecular defects leading to obesity & diabetes may be via several hierarchical genetic or epigenetic levels of control (Figure 8): (1) Abnormal acetylation/deacetylation of transcription factors resulting in anomalous gene expression (2) Mutation of transcription factors leading to aberrant gene expression (c) Atypical expression of transcription factors themselves resulting in unnatural gene expression.

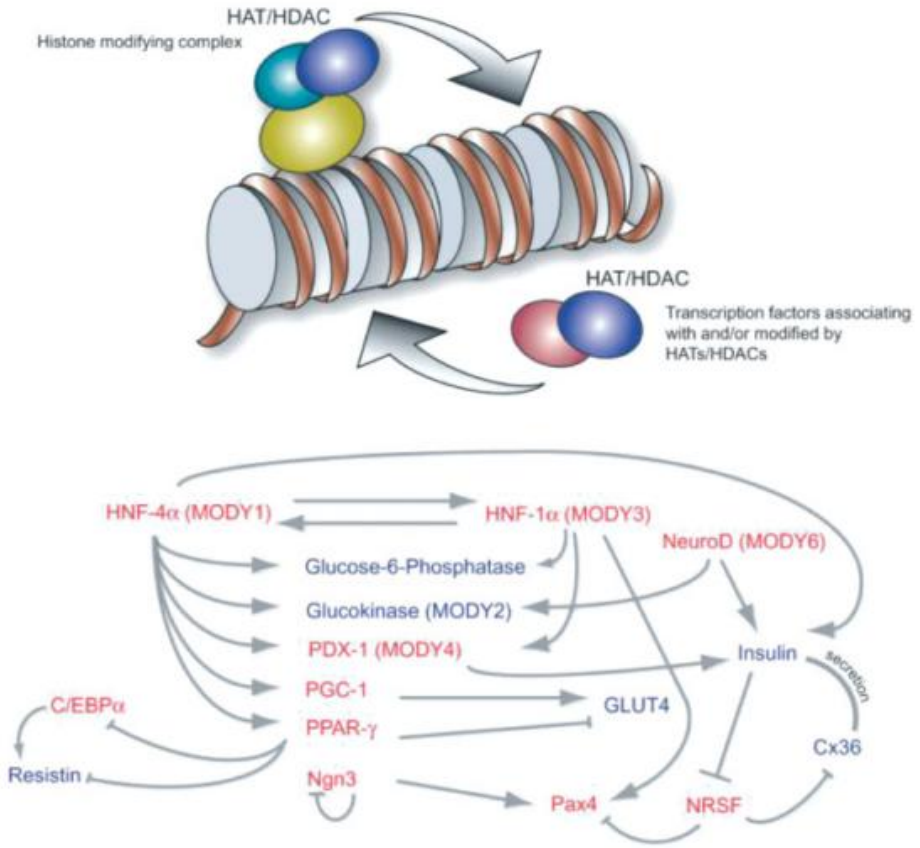


Figure 8. Schematic of how histone acetyltransferases and histone deacetylases may modulate the chromatin surrounding important diabetes genes either through direct interactions with specific transcription factors, or via specific complexes[104].

1.4.4. EPIGENETICS AND THE MATURE BETA-CELL

Pdx1 does not appear to have the intrinsic ability to either modify histones or directly alter chromatin structure, but once its transcriptional co-activators are recruited, they serve as a physical or functional bridge that allows Pdx1 to communicate with components of the basal transcriptional machinery, leading to either activation or repression of downstream target genes. For example, Pdx1 associates with HDAC1 and -2 under low glucose conditions (1-3mM) to repress *Insulin* gene expression, and recruits the transcriptional co-activator and HAT p300 to the *Insulin* promoter only in high glucose conditions (10-30mM)[161]. The HMT Set7/9 is recruited by Pdx1 to the *Insulin* and *Glut2* genes, leading to H3K4me2 and recruitment of RNA polymerase II[162]. Pdx1 modulates histone H4 acetylation to directly activate *Insulin* gene expression in β-cells[163]; exogeneous *Pdx1* expression in α-cells also leads to *Insulin* expression via H4 acetylation[164].

MODY is a monogenic autosomal form of T2DM, and 6 genes (HNF4α, GK, HNF1α, PDX1, HNF1α, and NEUROD1) that associate with HATs and HDACs have each been found to cause this disease. Mutations in PDX1 lead to MODY4, and as mentioned above, Pdx1 has a functional interaction with p300 that is required for *Insulin* gene transcription. Pdx1 also regulates its own transcription through histone H3 and H4 acetylation[165]. A mutant form of PDX1 in humans has an increased affinity for p300, but a lower DNA binding capability, suggesting that Pdx1 sequestration of HAT activity and its resultant chromatin remodeling patterns may be important in the development of diabetes in these MODY4 patients[166].

In IUGR rodents, mSin3A/HDAC is first recruited to the *Pdx1* promoter[167]; histone tails are deacetylated and *Pdx1* transcription repressed (Figure 9). At the neonatal stage, this epigenetic process is reversible and might define an important developmental window for therapeutic approaches. As dimethylated H3K9 accumulates, DNMT3A is recruited to the *Pdx1* promoter and initiates *de novo* DNA methylation, which locks in the *Pdx1*-silenced state in the IUGR

adult pancreas, resulting in diabetes[167]. *Pdx1* and *MafA* expression and activity are decreased in diet-induced obese (DIO) animals[168] and in cell culture conditions mimicking diabetes[169], leading to a decrease in insulin biosynthesis and secretion. A mutation in p300 abolishes binding to NeuroD1 and destroys the ability of p300 to activate *Insulin* E-box directed transcription in β-cells, leading to MODY6, thus showing the importance of p300 interactions with E-box activators in *Insulin* gene transcription[166]. Co-immunoprecipitation studies show that insulinoma-associated (IA)-1, cyclin D1, and HDACs functionally interact to regulate NeuroD1[170]. Small heterodimer partner (SHP) is one of the unique members of the orphan nuclear receptor superfamily that lacks a conventional DNA-binding domain[171]. SHP directly regulates the transcriptional activity of Hnf4α[172] and acts as a co-regulator of NeuroD1 via physical interaction, repressing it by competing with p300 for binding[173].

Therefore, the association of many β-cell transcriptional regulators with HATs, HMTs and HDACs provide correlative evidence for histone modification in regulating β-cell function or growth.

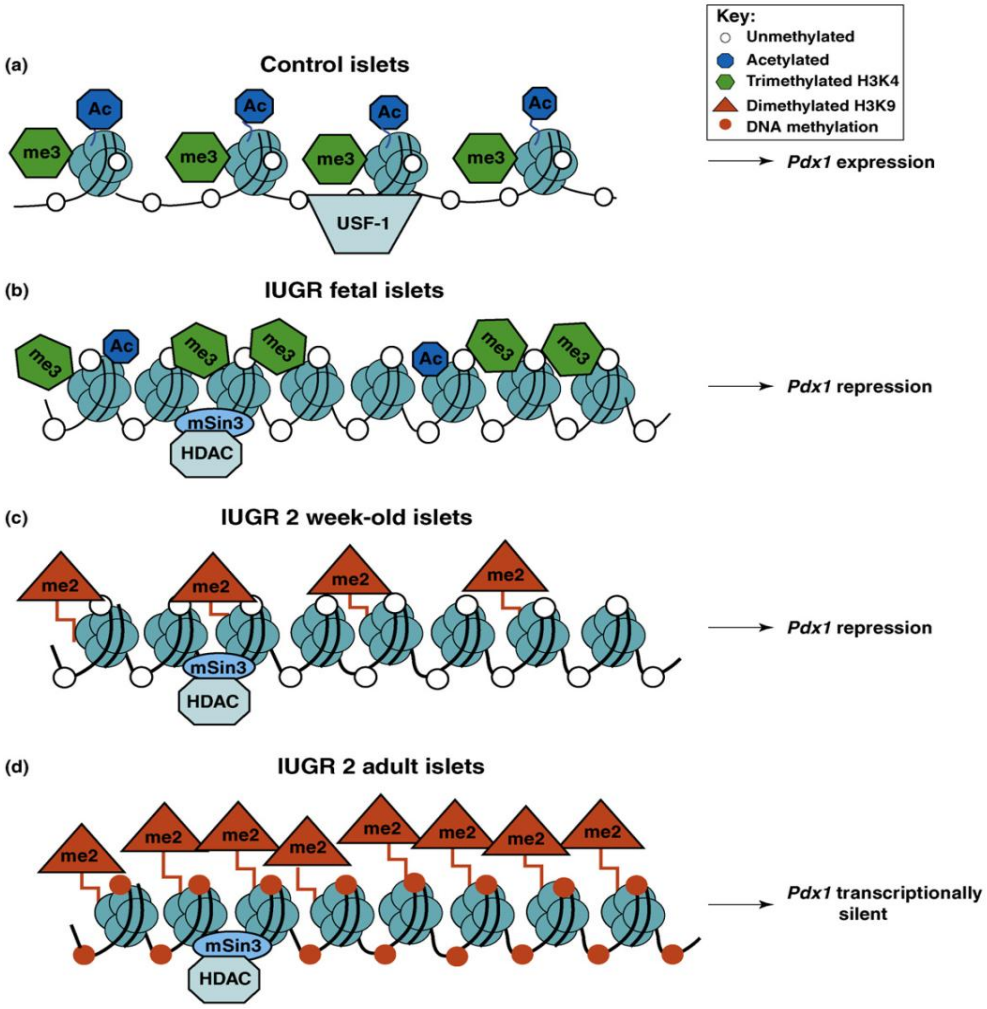


Figure 9. Epigenetic changes at *Pdx1* in IUGR rats during the development of T2DM[167]. (A) In pancreatic β-cells, the *Pdx1* proximal promoter is normally found in an unmethylated (white circles) open chromatin state allowing access to transcription factors such as Usf1 and is associated with nucleosomes characterized by acetylated (Ac, blue octagons) histones H3 and H4 and with trimethylated H3K4 (Me, green hexagons). (B) In IUGR fetal and (C) 2 week-old islets, histone acetylation is progressively lost through association with a mSin3A–HDAC1–DNMT1 repressor complex, with trimethylated H3K4 disappearing and dimethylated H3K9 (Me, red triangles) appearing after birth. (D) IUGR adult islets are characterized by inactive chromatin with dimethylated H3K9 and extensive DNA methylation (red circles) locking in the transcriptionally silent state of *Pdx1*.

1.4.5. GLOBAL EPIGENETIC PROFILING OF THE PANCREAS

Genes that have similar and/or co-ordinated expression patterns in specific tissues or at specific developmental stages are often clustered, like the *Hox* and *β-globin* loci[174]. The regulation of such gene clusters is often co-ordinated by shared common cis-regulatory elements, which activate expression through mechanisms that typically result in histone modifications associated with transcriptionally active chromatin. For example, genes in the *β-globin* cluster are activated by a shared locus control region, which, through chromatin loop formation, makes physical contact with individual promoters of the cluster in a developmental stage-specific manner[175]. The human INSULIN gene has recently been found to be part of a large open chromatin domain specific in islets which is marked by histone acetylation and H3K4 di-methylation marks distributed over the entire coding region and an 80-kb region around it[176]. This region also contains the {tyrosine hydroxylase (TH)-(INS)-insulin-like growth factor 2 antisense (IGF2AS)-insulin-like growth factor 2 (IGF2)} gene cluster[176] and their islet specific co-ordinate gene expression, suggesting their possible function in the maintenance of this unusually large region of open chromatin.

In another recent study that looked at open regions of chromatin in human pancreatic islets[177], the authors identified islet-specific clusters of open chromatin sites (termed COREs (clusters of open regulatory elements)), of which many genes encoding proteins with known function in islets (such as PDX1, PAX6, GLP1R, SLC30A8 and CPE), belong. They overlapped these CORE regions with 20 loci that have previously been associated with either T2DM or fasting glycemia, and found that 38 single nucleotide polymorphisms (SNPs) at 10 loci overlapped with islet-specific regions of open chromatin. This included the T2DM-associated SNP rs7903146 in the transcription factor 7–like 2 (TCF7L2) gene, which encodes a transcription factor involved in Wnt signaling and is a well-established risk locus for T2DM[178]. They detected an allelic imbalance in open chromatin signal in nine islet samples obtained from individuals who were heterozygous for the T-risk allele of rs7903146, indicating that the T allele is located in more open chromatin than is the non-risk C allele (Figure 10). This

suggests that the T allele may exert its effect on T2DM risk by altering the accessibility of the TCF7L2 DNA, and they showed in luciferase assays that the T allele had greater enhancer activity than the C allele[177].

In another new study, open chromatin regions, CCCTC-binding factor (CTCF)-binding sites, H3K4me3, H3K4me1, and H3K79me2 were profiled across the entire genome in human islets[179]. Integrated analysis of these large-scale data sets identified 18 000 putative TSSs, 30% of which were previously unannotated by RefSeq and at least several hundred of these are islet-active TSSs, including those for major islet miRNAs previously implicated in the control of glucose homeostasis[180]. Interestingly, active chromatin marks (H3K4me3 and H3K79me2) were absent from a subset of highly islet-expressed genes, including those encoding islet-specific hormones (INS, GCG, SST, IAPP, PPY, and Transthyretin (TTR). This observation suggests that some genes critical for islet function could have an unconventional promoter chromatin signature, indicative of a unique transcriptional control mechanism, and a similar proposal was also put forth by the group that found that the Insulin gene was part of a large open chromatin domain[176]. Fifty SNPs associated with islet-related diseases and traits were also found to map to within 500bp of a candidate non-promoter regulatory element, and 4 of the 12 elements that function as enhancers in vitro: FTO, Potassium voltage-gated channel, KQT-like subfamily, member (KCNQ)-1, TCF7L2, and Wolfram syndrome (WFS)-1 harbor T2D-associated SNPs, including two (TCF7L2 and WFS1) that exhibit significant allele-specific differences in activity[179]. These results suggest that altered enhancer activity plays a role in the molecular mechanism underlying at least a subset of T2D genetic association signals.

In summary, there is a growing body of current literature suggesting a role for epigenetic interactions in the complex interplay between gene regulation and the environment. Nevertheless, our knowledge about the molecular mechanisms linking environmental factors and T2DM still remains limited.

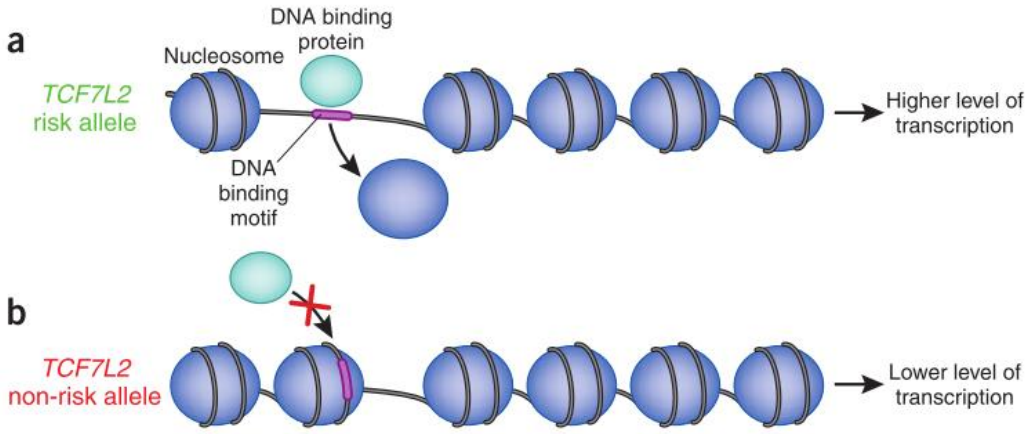


Figure 10. Model for the chromatin state and mechanism for the TCF7L2 association in Type 2 Diabetes[181]**.** (A) The T risk allele of the T2D-associated variant in the TCF7L2 gene (SNP rs7903146) is associated with open chromatin, which may facilitate protein-DNA binding and associate with increased levels of transcription relative to the non-risk allele. (B) The non-risk C allele is associated with closed chromatin, which may reduce access for protein binding.

1.5. EXPERIMENTAL HYPOTHESES AND APPROACH

Type 2 Diabetes arises from the inability of pancreatic β-cells to compensate for insulin resistance induced by environmental factors in genetically predisposed individuals[48]. Chronic hyperglycemia[182], hyperlipidemia[183], or a combination of both[184], have been proposed to contribute to β-cell failure. Major goals in diabetic research are to restore the functionality of damaged β-cells or to replace them. Although insulin replacement remains the mainstay of treatment in T2DM, considerable effort has been devoted to developing approaches for cell-based therapy. Through these efforts, it may become possible to introduce cells into diabetic patients that not only produce insulin, but also respond appropriately to metabolic signals that regulate insulin production, processing, and secretion. Accomplishing this goal requires a better understanding of the essential determinants of the β-cell phenotype at the physiological and molecular level. Pdx1, NeuroD1 and MafA are transcription factors necessary for maintaining mature β-cell function and critical for pancreas development[85]. Epigenetic changes to the chromosome architecture in β-cells have been associated with the loss of both glucose sensitivity and insulin production[130]. The use of genome-wide technologies to study gene expression and regulation in patients or animal models with T2DM has increased rapidly over recent years, generating long lists of new T2DM candidate genes. However, the use of global techniques to study epigenetic modifications in these systems has been limited, and epigenetic changes associated with T2DM are therefore still poorly understood. Therefore, understanding how these transcription factors regulate mature β-cell function, and the role of epigenetic changes associated with these transcription factors may provide useful knowledge in deriving β-cells from ESCs, in differentiating reprogrammed ESCs from diabetic patients, and in directing or blocking specific epigenetic changes in β-cells to restore functionality.

I have taken a two-pronged approach in my PhD dissertation, where my main hypotheses are:

1. The molecular mechanisms governing the dynamic regulation and/or co-regulation of Pdx1, NeuroD1 and MafA are key determinants of β-cell function.

2. As β-cell dysfunction develops, modulation of the chromatin environment adversely affects expression and activity of Pdx1, MafA, NeuroD1, and possibly other key β-cell transcriptional regulators.

To test these hypotheses, the aims are to:

1. Define the Pdx1, NeuroD1, and MafA transcriptional regulatory network operative in β-cells.

2. Establish the epigenetic profile of β-cells in order to investigate the role of chromatin remodeling in β-cell dysfunction.

For these purposes, I have used a combination of genome-wide sequencing technologies combined with gene expression analyses to profile the global transcriptional occupancy of Pdx1, NeuroD1 and MafA, thus correlating transcriptional occupancy with gene regulation. I have also functionalized some of these target genes in order to gain deeper insight into the Pdx1, NeuroD1 and MafA-regulated processes. I have profiled global epigenomic changes in pancreatic β-cells with an impaired GSIS response. By investigating the molecular mechanisms of β-cell regulation and dysregulation, it will provide new insights on (1) How these important transcription factors work to regulate mature β-cell function, (2) How the epigenome is linked to pancreatic transcription factor-mediated activity in GSIS and the epigenetic mechanisms associated with β-cell dysfunction, and (3) Provide clues as to how the chromatin environment may be modified as a potential therapeutic avenue in treatment of diabetes. In addition, this information will be useful in developing *in vitro* ESC differentiation protocols to produce functional glucose-responsive β-cells.

1.6. SIGNIFICANCE OF WORK

Some other pertinent questions are whether the epigenetic changes induced by today's sedentary lifestyle can be inherited by coming generations, and whether these changes are reversible. Clinical trials with HDAC inhibitors have repeatedly demonstrated low toxicity despite the expectation that they may cause global gene dysregulation, and preliminary studies indicate that they may prove useful in treating diabetes[185]. It is desirable that the balance of genes whose regulation would be altered with by epigenetic modifiers would tip the balance from unfavorable diabetogenes to those with enhanced benefits. With the insights gained from these experiments, it is hoped that eventually, genome-wide transcriptional and epigenetic characterization will lead to patient-specific therapies that also allow monitoring of genome-wide epigenetic consequences of these therapies.

CHAPTER 2:

METHODS & MATERIALS

2.1. CELL CULTURE

2.1.1. NIT-1 CELLS

NIT-1 mouse insulinoma cells (ATCC) were grown in Kaighn's modification of Ham's F12 media (F12K) (Life Technologies), with 1.5g/l sodium bicarbonate, supplemented with 10% dialyzed fetal bovine serum (Life Technologies). They were grown in a monolayer with a change of media every 2-3 days, and passaged at a 1:2-3 ratio when the dish was ~70% confluent, with the following procedure: Each dish was rinsed twice with 1xPBS solution (Gibco) and incubated in 3ml 0.05% Trypsin (Gibco) for 3 minutes at 37°C to allow detachment of cells from the culture surface. The trypsin-treated cells were diluted with 7ml NIT-1 culture media, pelleted in a centrifuge for 5 min/500rpm, and re-suspended in 5ml fresh NIT-1 media. The cells were titrated to achieve a single cell suspension and seeded.

2.1.2. MOUSE EMBRYONIC STEM CELLS (ESCS)

Feeder-free E14 mouse ES cells (ESCs) were maintained on 0.1% gelatin coated dishes in E14 proliferative medium containing DMEM/15% ES FBS (Gibco), 0.1mM MEM non-essential amino acids (Gibco), 2mM L-glutamine (Gibco), 0.1mM β-mercaptoethanol (Gibco), and CHO- Leukemia inhibitory factor (LIF) (1,000 U/ml). The cells were passaged every 2 days in a similar fashion as the NIT-1 cells, and seeded on gelatin-coated plates at a density of $4x10^6$ cells per $10cm^2$ dish. Mouse ES cells cultured on mouse embryonic fibroblast (MEF) feeder layers were maintained in the proliferative medium described above, with the exception of CHO-LIF.

2.1.3. DIRECTED ESC DIFFERENTIATION TOWARDS THE PANCREATIC LINEAGE

In stage 1, ESC differentiation was initiated by generating embryoid bodies, to allow cells to form a three-dimensional structure that allows intra-cellular signaling. LIF was withdrawn to promote spontaneous differentiation. In Stage 2, serum was withdrawn and Nodal signaling

was promoted with the addition of Activin A (R&D Biosystems) added to potentiate endoderm formation, FGF4 (R&D Biosystems) was added to induce anterior gut fate. In stage 3, Activin A and FGF4 were maintained in the culture medium and bone morphogenic protein (BMP) 4 (R&D Biosystems) 4 and FGF10 (R&D Biosystems) were added to expand the population of putative pancreatic progenitors. Pancreas specification depends on addition of RA and suppression of SHH by cyclopamine (Sigma), which represses the mesodermal markers Gsc, Bra. Low glucose was used in stages 4 and 5 because glucose restriction may initiate generation of insulin producing cells. Nicotinamide (Sigma), a SIRT-specific inhibitor, is commonly used as part of the differentiation cocktail to drive ESCs towards the β-cell lineage[186]. β-cellulin (R&D Biosystems), a glycoprotein highly expressed in the pancreas, was added to promote the differentiation, regeneration and proliferation of pancreatic β-cells.

2.1.4. Cryopreservation of Cell Lines

Confluent cell cultures were trypsinised as described above and centrifuged for 5 minutes at 500rpm. The cells were re-suspended in freezing media (70% DMEM with 20% FBS and 10% dimethyl sulphoxide) and transferred into Nunc cryotubes at 1ml per vial. The cryotubes were placed at -80°C overnight in Mr Frosty containers buffered with isopropyl alcohol and then transferred to a liquid nitrogen tank the following day for long-term storage.

2.1.5. Thawing of Cell Lines

Frozen cryotubes were retrieved from liquid nitrogen tanks and placed immediately into a 37°C water bath. The cells were rapidly thawed and then transferred into 10ml of pre-warmed culture media. The cell suspension was pelleted for 5 minutes at 500rpm, then gently resuspended in culture media and plated in 10 cm^2 tissue culture dishes to allow for overnight adhesion of cells to the vessel surface.

2.2. CELL-BASED ASSAYS

2.2.1. RNA INTERFERENCE (RNAI) ON NIT-1 CELLS

RNA interference (RNAi) experiments were performed with Dharmacon siGENOME SMARTpool reagents or Ambion Silencer siRNAs against mouse genes (Appendix 1).The Dharmacon si*CONTROL* nontargeting siRNA pool or the Ambion siNTC was used as a negative control. NIT-1 cells were transfected according to manufacturer's instructions in 12-well plates at a density of 2.5 x 10^5 cells/well or 6-well plates at a density of 4 x 10^5 cells/ well.

2.2.2. GSIS

NIT-1 cells were seeded in 12-well plates at a density of 4 x 10^5 cells and left to grow till 70% confluency before the experiment. The cells were rinsed 2X with PBS and starved for 2 hours at 37°C with RPMI (Life Technologies) before washing 3x with 1xKRH (10xKRH formulation: 15.194g NaCl, 700.8mg KCl, 326.6mg KH2PO4, 288.96mg MgSO4, to a volume of 200ml H_2O. 100ml 1xKRH: 10ml 10x KRH, 2ml 1M HEPES, 1% BSA, 256µl $CaCl_2$, pH 7.38-7.42). The cells were incubated in 500µl 1xKRH for 1 hour and the media removed and stored on ice (basal secretion). 500ul 1xKRH with 20mM glucose was then added and incubated for 1 hour and the media removed and stored on ice (stimulated secretion). The media from the basal and stimulated cells were spun at 500rcf for 5 minutes and the supernatant used for subsequent ELISAs.

2.2.3. GLUCOLIPOTOXICITY TREATMENT OF NIT-1 CELLS

150,000 cells/well were seeded in triplicate wells/ condition in a 12-well plate and grown to 60-70% confluence. Varying concentrations of 150mM palmitic acid (Sigma) and glucose (Sigma) was added to RPMI for 2 days.

2.2.4. ESTABLISHMENT OF THE PDX1-CLONAL MOUSE ESC LINE

The Ainv18 cell line (gift of George Daley, Harvard Medical School, Boston, MA) is an E14 mouse ES line modified for targeted gene insertion downstream of a doxycycline-responsive promoter 183. The *Pdx1* transgene was PCR-amplified from a mouse cDNA library and subcloned into the KpnI/XbaI sites of the pLox-N-tag-HA vector (George Daley, Harvard Medical School, Boston, MA) with the following primers, restriction sites indicated in uppercase: Forward 5'-ATTACTCTCGAGTAACAGTGAGGAGCAGTA -3' and Reverse 5'-CGTACGGTCGACTCA CCG GGG TTC CTG CGG -3'. An additional nucleotide was inserted in the forward primer to allow in-frame processing of the *Pdx1* insert with the N-terminal HA tags. For targeted insertion at the HPRT locus, 20µg each of pLox-N-tag-HA-Pdx1 and pSALK-Cre were transfected with Lipofectamine 2000 (Life Technologies) into $6x10^5$ AINV18 cells seeded in $10cm^2$ dishes. For derivation of Ainv18-*Pdx1* clonal lines, mouse ESC medium was supplemented with 350µg/ml G418 solution (Life Technologies) 24 hours after transfection and maintained for 14 days. Colonies arising from G418 selection were individually picked and expanded on neomycin-resistant MEFs. Site-specific integration was confirmed by PCR analysis with the following primers: LoxinF 5'-cta-gat-ctc-gaa-gga-tct-gga-g-3' and LoxinR 5'–ata-ctt-tct-cgg-cag-gag-ca-3'. To induce over-expression, AINV18-*Pdx1* cells were treated with 1µg/mL doxycycline in mouse ESC medium. Protein was harvested to test and confirm the efficacy of regulation by doxycycline.

2.3. MOLECULAR BIOLOGY TECHNIQUES

2.3.1. CLONING

The sub-cloning of DNA fragments into vectors was achieved by restriction enzyme digestion, gel purification of digested DNA, and T4 ligase treatment to yield the final construct. Restriction digests were preformed overnight at 37^oC according to manufacturers' instructions (New England Biolabs). The following day, digested DNA was resolved on 1-1.5% agarose gels in TAE buffer, and the desired fragments were visualized under UV

lighting and excised from the gel. The DNA fragments were released from agarose matrix using the Qiagen gel extraction kit according to manufacturer's instructions. The concentration of the recovered DNA was quantified by UV absorbance using the nanodrop instrument, and ligations were performed at 5:1 (insert:vector) molar ratio in 20µl reactions containing 1U of T4 DNA ligase in 1xT4 ligase buffer (Roche) at 16^{o}C overnight.

2.3.2. TRANSFORMATION OF CHEMICALLY COMPETENT CELLS

For transformation of ligated plasmids, a 50µl aliquot of One-Shot Top10 chemically competent *E.coli* cells (Invitrogen) was thawed on ice. Approximately 10ng of ligation reaction was added to the thawed cells and gently mixed followed by incubation on ice for 30 minutes. Transformation was induced by heat-shock treatment at 42ºC for 1 minute. The cells were cooled on ice for 2 minutes, and 200µl of nutrient-rich LB-SOC was then added to the tube for incubation at 37ºC for 1 hour in a bacterial shaker. Following incubation, 50-300ul of the culture was plated on LB-agar plates containing the appropriate antibiotic selection, and the plates were placed in an incubator at 37ºC overnight to allow growth of bacterial colonies.

2.3.3. ISOLATION OF PLASMID DNA FROM BACTERIA

Bacterial colonies containing positive ligation events were expanded as follows: 5ml LB medium with appropriate antibiotic selection was inoculated for overnight culture at 37ºC. The overnight cell suspensions were centrifuged at max speed for 15 minutes to pellet bacteria. The bacterial cells were then lysed and plasmid DNA purified using the Qiagen miniprep kit according to manufacturer's instructions. The DNA was eluted in 50µl dH_2O and quantified by Nanodrop to ensure a final concentration in the range of 200 – 1000 µg/ul.

2.3.4. PREPARATION OF BACTERIAL STOCKS

Bacterial stocks of successful transformation events were prepared in 33.3% sterile glycerol. The LB-glycerol suspension was vortexed to ensure thorough mixing, cooled on ice for 1 hour, and then placed at -80^{o}C for long term storage.

2.4. GENE EXPRESSION ANALYSIS

2.4.1. RNA EXTRACTION

RNA was extracted from mammalian cells with TriZol reagent (Invitrogen) and then further purified with the RNeasy minikit (Qiagen) according to manufacturer's instructions. RNA samples were treated with DNase for removal of genomic DNA contamination. This was performed in 100 µl reactions containing 5 µg RNA, 0.2 U/µl RNAse inhibitor, and 0.15 U/µl RNase-free DNase (ProMega). Samples were incubated at 37^{o}C for 30 minutes, followed by heat-inactivation of enzymes at 65^{o}C for 15 minutes. The RNA samples were then purified using two phenol extractions and one chloroform extraction, and back extractions were performed with each step using equal volumes of sterile dH_20 to minimize loss of RNA. Total RNA was precipitated, washed with 70% ethanol and finally re-suspended in 30µl of sterile nuclease-free dH_20.

2.4.2. CDNA SYNTHESIS

cDNA was synthesized with 100ng - 1 µg total RNA using the High Capacity cDNA Archive kit (Applied Biosystems) as per manufacturer's instructions. Briefly, cDNA synthesis was performed in 20µl reactions consisting of RNA diluted in 10 µl nuclease-free dH_20, and cDNA Archive kit components: 1xReverse Transcriptase buffer, 5.5mM $MgCl_2$, 500µM dNTP mix, 2.5µM random primers, 0.25U/µl MultiScribe Reverse Transcriptase. Synthesis reactions were incubated in a PCR thermocycler at 25^{o}C for 10 minutes followed by 37^{o}C for 2 hours. The cDNA samples were stored at -20^{o}C following synthesis.

2.4.3. QUANTITATIVE REAL-TIME PCR

Quantitative PCR assays were performed as follows: Reversed transcribed cDNA samples were diluted 10x in nuclease-free water, and added to 5µl TaqMan® Universal PCR Master Mix reagent (Applied Biosystems) containing 0.5µl of a single TaqMan probe (20xTaqMan® Gene Expression Assay reagents; Applied Biosystems). The list of TaqMan probes are provided in Appendix 2. Reactions were conducted in duplicate within 384-well reaction plates (ABI) at a final volume of 10µl on the ABI Prism 7900 machine. For SYBR Green reactions, primers against the gene of interest were designed with the following criteria: spanning at least one intron, 15-25 bp in length with a 40-60% GC content, and a melting temperature™ of 57-63^{o}C.

2.4.4. ILLUMINA MOUSE ARRAYS

The whole-genome Illumina Mouse Ref-8 v2.1 beadchips were used for microarray analysis of RNAi-treated samples. These Illumina beadchips allow interrogation of approximately 25,600 well annotated mouse Refseq transcripts. For the microarray analyses, mRNA was harvested from 3 or 4 biological replicates per experiment. These samples were reverse-transcribed, Cy3-labeled, and hybridized to Illumina MouseRef-8 v2.1 Expression Beadchips according to the manufacturer's protocol (Illumina Inc.). Following the wash of Illumina chips, microarray signals were scanned with an Illumina BeadStation array reader, and the data were background-normalized and assessed for quality with cluster dendrograms and DirectHyb control plots (Beadstudio v1.5, Illumina).

2.4.5. Statistical Analysis of Microarray Data

The microarray data were processed with Genespring GX v10.2 (Agilent Technologies). The unpaired Student's *t* test was used to determine statistical significance with Benjamini and Hochberg multiple-testing correction (FDR<0.05), and genes demonstrating ≥ 1.5-fold change were reported as statistically significant. For statistical analysis of overlapping genes between RNAi samples, computer simulations were conducted as follows: a number of genes corresponding to the actual number of genes within each list were randomly sampled from the Illumina mouse Ref-8 v2.1 Beadchip. The random gene lists were overlapped to derive a percentage, and this process was repeated 100 times to yield an average percentage overlap by chance. Finally, *p*-values were computed by comparing the 100 simulated overlaps to the actual overlap percentage using one-sample *t*-tests.

2.4.6. Functional Annotations Using the Panther Database

The microarray gene lists were uploaded to the Panther database for functional annotations (http://www.pantherdb.org). In this analysis, the Panther Gene Expression tool set was used to classify differentially-regulated genes under biological process groups using the "Compare Gene List" function, relative to the 25,600 Illumina mouse Ref-8 v2.1 transcripts, and the binomial test was used to identify statistically significant over- or under-represented functional categories within the RNAi or over-expression gene lists.

2.4.7. Functional Annotations Using Ingenuity Pathway Analysis

The RefSeq list of genes was uploaded into IPA (Ingenuity® Systems)[187] and identified the top biological networks and pathways associated with the transcription factor binding sites for Pdx1, NeuroD1 and MafA. The same procedure was performed with the peaks from the H3K4me3, H3K27me3 and FAIRE libraries. Each identifier was mapped to its corresponding object in the Ingenuity® Knowledge Base and termed network eligible molecules. These were

then overlaid onto a global molecular network developed from information contained in the Ingenuity Knowledge Base, and networks were then algorithmically generated based on their connectivity. The functional analysis identified the biological functions and/or diseases that were most significantly enriched to the data set.

2.5. PROTEIN EXPRESSION ANALYSIS

2.5.1. CELL LYSIS AND PROTEIN QUANTITATION

Confluent cell cultures were harvested by scraping and washing with 1xPBS solution. The cells were centrifuged for 5 minutes at 500rpm, and then re-suspended in cell lysis buffer comprising 50mM Tris pH 7.4, 150mM NaCl, 2mM EDTA, 1% NP-40, 0.1% SDS, and 1xEDTA-free Protease inhibitor (Roche). The cell lysate was incubated with rotation for 10 minutes at 4^{o}C and was cleared by centrifugation at 12,000 rpm for 30 minutes at 4^{o}C. The supernatant was transferred to a fresh tube and protein was quantitated using the Bradford assay (Bio-Rad Laboratories). Six bovine serum albumin (BSA) protein standards were used at 0µg, 50µg, 100µg, 200µg, 300µg and 400µg for the Bradford assay. Protein samples and standards were diluted in 0.15M NaCl prior to triplicate loadings onto a 96-well ELISA plate. Absorbance values were measured on a Sunrise Tecan Microplate Reader at A595 nm.

2.5.2. SDS-PAGE

Denaturing polyacrylamide gel electrophoresis was conducted to resolve proteins by size. Thirty micrograms of protein was added to 6x SDS loading buffer (250mM Tris/HCl pH6.8, 30% glycerol, 10% SDS, 5% β-mercaptoethanol, and 0.02% bromophenol blue), followed by incubation at 95^{o}C for 5 minutes. Protein samples were loaded onto a 10% or 12% acrylamide gel for electrophoresis at 100V for 1.5 hours. The resolved protein samples were then electroblotted onto Hybond C extra nitrocellulose membrane (Amersham) for 1 hour in Western Transfer Buffer (20% Ethanol, 70% dH_20, and 1xWestern Transfer Buffer stock

containing 14.5g Glycine and 3.0g Tris base). Consistency of protein loading and quality of Western blot transfer was assessed by Ponceau S staining (0.5% Ponceau S, 1% glacial acetic acid). In this protocol, the electro-blotted membrane was rinsed once in Tris-buffered saline containing Tween-20 (TBS-T; 10mM Tris/HCl pH 8.0, 150mM NaCl, 0.5% Tween-20; pH 8.). Ponceau S was then applied for 30 seconds to enable visualization of proteins on the membrane, following which the membrane was rinsed twice in dH_20 and then thrice in TBS-T. The nitrocellulose membranes were immediately probed for protein or stored at 4^oC in air-tight boxes containing TBS-T.

2.5.3. Protein Detection and Chemiluminescence Detection

Immunodetection of proteins was carried out as follows: Nitrocellulose membranes were blocked for 1 hour at room temperature in 5% (w/v) skimmed milk powder and 1% BSA. Protein detection was performed with rabbit-anti-Pdx1 antibody (1:500; Upstate) or rabbit-anti-MafA (1:500, Abcam), or goat-anti-NeuroD1, sc:590, Santa Cruz Biotechnology), followed by HRP-conjugated donkey-anti-mouse or donkey-anti-rabbit secondary antibodies (Santa Cruz Biotechnology) at a dilution of 1:5000. Loading consistency was determined with mouse-anti-β-actin or donkey anti-GAPDH (1:1000; Invitrogen) and goat-anti-donkey HRP secondary antibodies (1:5000; Santa Cruz Biotechnology). The primary antibodies were incubated overnight at 4^oC and the secondary antibodies were incubated for 1 hour at room temperature, with 3xTBS-T rinses for 15 minutes performed in between incubations. Chemiluminescence detection was carried out with the Enhanced-Chemiluminescence (ECL) Western blotting kit (Amersham) according to manufacturer's instructions.

2.5.4. Enzyme-Linked Immunosorbent Assay (ELISA)

The amount of insulin was measured quantitatively using the Mouse Insulin ELISA (Mercodia) per manufacturer's instructions. Briefly, the samples were diluted 20x and 10µl of samples/standard loaded in each well, with 100µl of enzyme conjugate media. The plate was

shaken at 21°C, 900rpm for 2 hours, then washed 6x with wash buffer. 200µl of TMB was added to the wells and incubated for 15 min and the reaction was stopped with 50µl stop solution. Insulin in the sample binds with (HRP)-conjugated anti-insulin antibody as well as an anti-insulin antibody is bound to the wells of a 96-well plate, and the bound HRP complex is detected by a reaction with TMB and the output measured with the Sunrise Tecan Microplate Reader at wavelength 450nm. The standards were averaged and a scatter plot point to point graph was plotted. The OD of the samples was averaged. Depending on the OD of the samples, the appropriate formula was used to calculate the insulin concentration. The secreted insulin concentration (S) is divided by the basal insulin concentration (B), to obtain the percentage increase of secreted insulin; % increase==S/B*100%. The triplicate % increase was averaged. The standard deviation and error bars were calculated with the triplicate % increases.

2.5.5. CHROMATIN IMMUNOPRECIPITATION (CHIP)

NIT-1 cells were cultured to a density of $1x10^8$ cells for each IP. Two biological ChIP replicates were performed per experiment. Cells were cross-linked for 10 minutes at room temperature with 1% (w/v) formaldehyde and the reaction was subsequently quenched with 125mM glycine. Nuclear fractions were isolated with a 3-step protocol that involved washing with increasing concentrations of SDS and high-speed centrifugation, and the cross-linked DNA was sheared to lengths of 200-500 bp with a Branson water sonicator. Chromatin was pre-cleared with thrice-washed Dynabeads for an hour at 4°C, and 6µg of ChIP antibody (Pdx1 (Upstate), NeuroD1 (Santa Cruz), MafA (Abcam), H3K4me3, H3K27me3) was incubated with the pre-cleared chromatin overnight. Dynabeads (Life Technologies) were added to the antibody-chromatin complexes the next day and incubated for 3 hours at 4°C. Following incubation, the antibody-bead-chromatin complexes were rinsed and pelleted, and re-suspended in fresh 0.1% ChIP buffer (50mM HEPES, 150mM NaCl, 2mM EDTA, 1% Triton x-100, 0.1% NaDOC, 0.1% SDS), and washed 2 times. The antibody-bead-chromatin complexes were then washed with a series of increasingly stringent buffers and TE before

eluting at 68°C, 1400rpm for 40 minutes. Cross-links were reversed and protein removed by incubating the chromatin at 42°C for 2 hours with Pronase (0.2mg/ml) and an overnight incubation at 68°C. Finally, the samples were treated twice with phenol/chloroform/Isoamyl alcohol (25:24:1), then precipitated and rinsed once in 70% ethanol before re-suspension in 10-50µl of nuclease-free water.

2.5.6. Quantitative PCR for ChIP Enrichment

Quantitative PCR (qPCR) for ChIP enrichment was performed on the ABI PRISM 7900 machine with 1xSYBR Green PCR mastermix (Applied Biosystems), under the following conditions: 50°C for 2 minutes, 95°C for 10 minutes, and 40 cycles of 95°C for 15s and 60°C for 1 minute. Fluorescence from the amplified products was measured by laser spectral analyses on the ABI Prism 7900 machine, and the data were processed and displayed in the form of real-time amplification plots generated by the SDS software (v2.2; Applied Biosystems). ChIP fold-enrichment was determined by normalizing Threshold cycle (Ct) values of ChIP samples against sonicated whole cell DNA extract, and then subsequently to a non-enriched ChIP control region set at a value of 1. Details of primers are provided in Appendix 3. All PCR primers gave a single product as confirmed by heat dissociation analysis on the ABI Prism 7900 (Applied Biosystems), under the following conditions: 95°C for 15s, 60°C for 20s, 20 minute ramp to 95°C and 95°C for 15s. The data from this analysis were compiled by the SDS software (v2.2; Applied Biosystems) and displayed as melting curve graphs. The graphs were checked for specificity of PCR amplified product as indicated by a single, well-defined melting curve peak.

2.5.7. Formaldehyde-Associated Isolation of Regulatory Elements (FAIRE)

NIT-1 cells were harvested for Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) [188]. Briefly, the cells were washed 2x with PBS and 1% formaldehyde was added to

~70% confluent 10-cm dishes of NIT-1 cells for 10 minutes - this was the FAIRE sample. The cross-linked chromatin was sheared to a range of 100-500bp and cross-links were reversed by incubating an overnight incubation at 65^{o}C. The DNA was extracted in a double phenol/chloroform/Isoamyl alcohol (25:24:1) step, then precipitated and rinsed once in 70% ethanol before resuspension in 10-50µl of nuclease-free water. The FAIRE reference library was prepared in a similar fashion, except the cells were not fixed with formaldehyde.

2.6. Bioinformatics Analysis of ChIP and FAIRE Libraries

2.6.1. ChIP- and FAIRE-Sequencing Library Generation and Identification of Enriched Regions

10ng of purified ChIP or FAIRE sample (determined by Picogreen) was prepared for deep sequencing according to manufacturer's protocols (Illumina, Inc). Briefly, the DNA was end-repaired and blunt ended fragments generated. "A" bases were added to the 3' end of the DNA fragments, leaving a 3'-dA overhang. Adapters were ligated to these overhangs and PCR enrichment of adapter-modified DNA fragments was performed. The samples were run on an agarose gel and 200-300bp fragments excised and purified. The amplified library was assayed for quantity and quality with the Agilent Bioanalyzer 2000, and 5nM submitted for sequencing.

ChIP and FAIRE libraries were prepared in biological replicates and the raw reads from each biological replicate was pooled together and combined unique reads were filtered and used for subsequent analyses. HOMER[189] and MACS algorithms were used to perform peak calling, binding peaks were defined as those above the FDR (False Discovery rate) cutoff of 0.001 and poisson p-value ~ 2.7^{-07}. Transcription factor MACS or HOMER peaks were overlapped to the University of California, Santa Cruz (UCSC) genome data browser, using the Feburary 2006 (NCBI36/mm8) *Mus musculus* reference assembly.

In the epigenetic analysis, the raw mapped reads (in BED format) for H3K4me3 and H3K27me3 were used for peak calling by HOMER[189]. Control samples (genome sequencing without antibodies for H3K4me and H3K27me) were used as the background input. Two-fold enrichment relative to the background and an FDR of 0.001 were used as the thresholds. For the FAIRE experiments, F-Seq[190] was used for peak identification based on the raw mapped reads. The sequencing output of DNA without formaldehyde fixation was used for control. The identified peaks from the control sequencing were subtracted from those from the FAIRE sequencing. The standard deviation of 6.0 was used for the threshold. The gene-wise distribution of the peaks was examined by means of the 'annotatePeaks' command of HOMER. The closest transcription start site from each peak was identified for gene annotation. For motif finding, the motif lengths of 8, 10, and 12bp were searched against the repeat-masked mouse genome (mm8). Position-specific weight matrices were based on the HOMER's database.

2.6.2. Association of Transcription Factor Enriched Regions in Genes

Enriched regions were associated with genes by idenfiying the closest annotated RefSeq gene within 10kb. Enriched regions >10 kb from any annotated RefSeq gene were not associated with any gene. Enriched regions were associated with only the closest gene. While some binding sites may regulate several neighboring genes, this is likely not the general case, and, while the approach described would miss such interactions, it eliminates a large number of incorrect associations. Sites were considered to be in promoter regions if they fell within 3 kb of a RefSeq TSS, within a enhancer if they fell between 2 kb and 50 kb from a RefSeq TSS, and distal if they were >50 kb from a RefSeq TSS. To identify overlapping and co-occupied sites, locations of sites of maximal enrichment were compared for Pdx1, NeuroD1 and MafA and resized to ± 65 bp from the peak center and overlapped by a single base pair.

2.6.3. De Novo Motif Analysis

To identify motifs associated with the ChIP-Seq libraries, a *de novo* motif discovery strategy was employed. ChIP-Seq peaks of Pdx1, NeuroD1 and MafA were input into HOMER, which extracted peak sequences with less than 70% repeats and accounted for GC biases by using randomly selected background sequences matching GC characteristics of input sequences

2.6.4. Overlapping of Histone Marks with Transcription Factor Binding Sites (TFBS)

Transcription-factor binding sites (TFBSs) were defined as the above-identified peaks for each TF. The surrounding regions (5kb upstream and downstream) of each TFBS were examined for the tag distribution of H3K4me3 and H3K27me3. As for the co-occupancy of the TFs, the TFBSs that overlapped within 50bp were merged into a single peak. The numbers of H3K4me3 and H3K27me3 tags were counted according to the distance from the center of the TFBS and then log10 transformed for modification level. The minimum modification level for H3K4me3 and H3K27me3 among the three TFBSs and co-occupied sites was taken and used for the basal level. Each modification level was normalized based on the basal level of H3K4me3 and H3K27me3 respectively. Associated genes were identified based on the location of the TFBS. The expression level of the genes that contain a TFBS with the H3K27me3 level increased by 1.5-fold in the dysfunctional cells was compared between the normal and dysfunctional cells.

CHAPTER 3:

TRANSCRIPTIONAL REGULATION IN PANCREATIC BETA-CELLS

3.1. INTRODUCTION

The highly specialized function of pancreatic β-cells is dependent upon sustained expression of important transcriptional regulators. To expand our knowledge of the β-cell transcriptional regulatory network and to explore what other genes are important in pancreatic β-cell function, I have chosen to work on Pdx1, NeuroD1 and MafA and their possible co-regulatory roles in modulating pancreatic β-cell function. These core transcription factors contribute to the hallmark characteristics of pancreatic β-cells by: (1) Activation of target genes that encode Insulin and other important β-cell transcriptional regulators, and (2) Recruiting co-activators that are necessary to mediate the above-mentioned regulation. In pancreatic β-cells, Pdx1, NeuroD1 and MafA have been shown to respectively occupy the A3, E1 and C1 elements within a 340-bp region upstream of the transcriptional start site of the *Ins* gene[191]. The *Ins* promoter is organized in a complex arrangement of discrete *cis*-acting sequence motifs, which serve as binding sites for ubiquitous and β-cell-specific transcription factors[192]. The co-ordinated interaction between these *cis*-sequence elements and *trans*-acting transcription factors at the promoter contributes to the β-cell-specific expression of the *Insulin* gene, and the exquisitely-controlled regulation of its expression in response to nutrients (e.g. glucose, hormone signalling, calcium levels and nutrient availability)[192,193].

Pdx1, a member of the large family of HD-containing proteins, is expressed in precursors of the endocrine and exocrine compartments of the pancreas and is essential for pancreas development, β-cell differentiation and maintenance of the mature β-cell[85]. In mature β-cells, Pdx1 transactivates the gene encoding *Insulin* and other genes involved in glucose sensing and metabolism such as *Glut2* and *Gk*[92]. *Pdx1*$^{+/-}$ mice are glucose intolerant, with increased islet apoptosis, decreased islet mass, and abnormal islet architecture, indicating that proper gene dosage of *Pdx1* is crucial for normal glucose homeostasis[194]. These findings are consistent with the report that humans heterozygous for an inactivating mutation of Pdx1 suffer from MODY4[195]. The 283 amino acid long Pdx1 protein is very homologous among different species. The N-terminus contains the transactivation domain. The middle region contains a

homeodomain, which is responsible for DNA binding and protein-protein interactions as a transcriptional activation mechanism. The role of the C-terminus is poorly understood[196].

Over-expression and targeted disruption studies have established a crucial role for NeuroD1 in regulating brain and pancreatic islet development[197]. NeuroD1 is expressed in pancreatic and intestinal endocrine cells and neural tissues and the *NeuroD1*-null mouse exhibits a mild reduction of δ-cells, a moderate loss of α-cells, and a severe reduction of β-cells[71]. This profound effect on β-cell number signifies a critical role for NeuroD1 in addition to *Insulin* gene transactivation in the β-cell, because insulin is not required for the β-cell formation or maintenance. NeuroD1 does not bind directly to DNA but binds as a heterodimer with Class A bHLH proteins, such as E12, E47, or HEB, to E-boxes within the *Insulin* promoter[198].

The bLZ transcription factor MafA binds to the C1 element of the *Insulin* promoter in β-cells[199]. During pancreas development, MafA is first detected at the beginning of the principal phase of insulin-producing cell production, comparatively later than Pdx1 and NeuroD, which are expressed from the early stage of pancreas development. In addition, while both Pdx1 and NeuroD1 are expressed in various cell types in the pancreas, MafA is expressed only in β-cells. Therefore, it is likely that MafA is the principal factor required for β-cell formation and function. *MafA*$^{-/-}$ mice display glucose intolerance and developed T2DM, with age-dependent pancreatic islet abnormalities[69]. Expression of *Insulin1*, *Insulin2*, *Pdx1*, *NeuroD1*, *Glut2* is decreased, and glucose-, arginine-, and KCl-stimulated insulin secretion was severely impaired in these animals.

An earlier chromatin immunoprecipitation-chip (ChIP-chip) study has shown that Pdx1 and NeuroD1 co-occupy 440 genomic loci[86], suggesting that a significant fraction of Pdx1 and NeuroD1 binding sites exist within regulatory modules, where they may control gene expression in a synergistic manner. However, it remains unclear whether MafA interacts at similar loci to co-regulate genes important for β-cell function, as no genome-wide binding

studies have been performed for MafA. It is also unclear what the downstream effectors of these transcription factors are, and how they contribute to maintaining the glucose-responsive state of β-cells. It is therefore important to elucidate the transcriptional networks regulated by Pdx1, NeuroD1 and MafA, where detailed knowledge of these pathways remain key to understanding mature β-cell function.

Hypothesis: The molecular mechanisms governing the dynamic and/or co-regulation of Pdx1, NeuroD1 and MafA are key determinants of β-cell function.

Questions I will address in this chapter are:

1) Where on the genome do Pdx1, NeuroD1 and MafA bind in β-cells? Is there evidence for co-regulation of these transcription factors at other genomic loci, apart from the *Insulin* gene?

2) Which genes are affected upon depletion of these transcription factors, and what biological network are they significantly involved in regulating?

3) How does transcription factor occupancy correlate with gene regulation in a genome-wide manner?

Currently, there are no suitable human β-cell lines available, as most of the β-cell lines derived are difficult to be propagated *in vitro*, and such lines gradually lose their differentiated function over the time *in vitro*[200]. Among all of the engineered human β-cell lines, only two cell lines β-lox5[201] and NAKT-15[202] continue to possess the normal β-cell characteristics and function to at least some degree, and only a few articles have been published using them. It is exceedingly difficult to immortalize human β-cells because of their short half-life (50 days). Their tendency to undergo apoptosis after islet isolation, combined with the loss of cell-to-matrix interaction and the stress induced by tissue culturing procedures renders the β-cells to become un-differentiated and undergo further apoptosis[203]. Due to the lack of availability of a

human β-cell line, despite the differences in the physiological and metabolic regulation of rodents and humans, I elected a rodent model for the basis for my experiments, because they still remain a good tool for studying known physiological mechanisms and discovering novel pathways of regulation[204].

There are several commonly studied mouse and rat β-cell lines, such as MIN-6 and NIT-1 (mouse), and INS-1 (rat). Although INS-1 cells are responsive to glucose within the physiological range, they require β-mercaptoethanol in the culture medium, which is toxic to the cells and irreversibly denatures their proteins[205]. Moreover, the genomic (databases, annotations and other scientists experimental data) and molecular biological tools and reagents (antibodies, siRNAs, etcetera) available for mouse are more readily available, so I elected to use a mouse β-cell line. MIN-6 cells express Glut2 and Gk and respond to glucose within the physiological range of glucose in the presence of nicotinamide[206], but oftentimes a sudden loss of GSIS from MIN-6 cells during the course of the passages is noticed, possibly due to an outgrowth of cells with a poor response to glucose or a reduced expression of the genes responsible for GSIS[207]. NIT-1 cells have high levels of *Insulin* mRNA with less than 5% containing glucagon, and none contain PP or Sst[208]. They robustly secrete insulin in response to glucose stimulation, though not at physiological levels of glucose. Since the aims of my thesis focus mainly on understanding the transcriptional and epigenetic regulatory network in pancreatic β-cells and do not include studying the effect of anti-diabetic drugs on the cells or involve β-cell transplantation into diabetic animals, I elected to use NIT-1 cells in my experiments.

I addressed my experimental questions using ChIP coupled with high-throughput sequencing (ChIP-Seq) technologies to map the Pdx1, NeuroD1 and MafA binding sites in a mouse insulinoma β-cell line NIT-1, and siRNA-depleted these transcription factors in order to look at how binding of these transcription factors correlate with the genes they regulate.

3.2. RESULTS

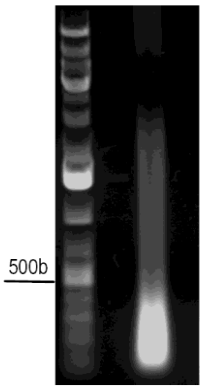
3.2.1. ChIP of Pdx1, NeuroD1 and MafA in NIT-1 cells

To test my hypothesis that Pdx1, NeuroD1 and MafA were bound at similar and dissimilar sites, I performed ChIP with chromatin isolated from the β-cell line NIT-1, with Pdx1, NeuroD1 and MafA antibodies, with rabbit and goat IgG antibodies as a control for non-specific protein-DNA interactions. Input DNA (sheared chromatin not used in any antibody pull-down) served as a control to measure genomic background, cell line variation and PCR-amplification bias. Chromatin was purified and sheared by sonication to obtain DNA fragments within a 100-500bp range (Figure 11A). To test the quality of the ChIP for Pdx1, NeuroD1 and MafA, I measured the signal obtained on the best characterized co-regulated gene, *Insulin*. Figure 11B shows the specificity of the Pdx1 antibody, as there was a 17-fold enrichment for the *Insulin* promoter using the antibody, versus a 1.2-fold enrichment with the non-specific rabbit IgG antiserum. Fold enrichment represents the abundance of enriched DNA fragments over a control region not bound by the antibody of interest, and many factors can affect it, including binding affinity of antibody, accessibility of target region, quality of the antibody-antigen interaction, the concentration of antibody, and the fold enrichment depends on both the absolute IP efficiency and on the background[209]. Pdx1 also bound to another region on the *Insulin* gene promoter (27-fold enrichment), and *Mtmr7* and *Pde43*. NeuroD1 and MafA were also bound to the same regions as Pdx1 (Figure 11C, D) with slightly different binding affinities (50-70 fold enrichment for NeuroD1 for the *Insulin* gene vs. 20-50 fold enrichment for MafA).

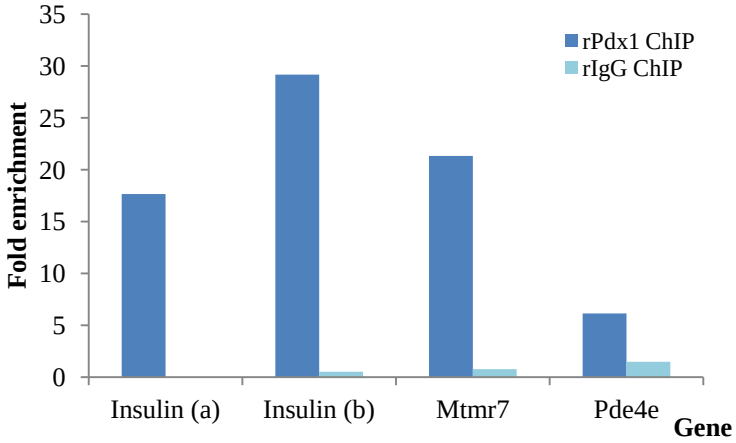
As mentioned earlier, a ChIP-chip paper had identified 440 sites in which Pdx1, NeuroD1 and MafA were found to co-occupy. I asked whether these sites were similarly occupied in my cells, and whether MafA was also occupying those sites. I selected 19 sites positively identified to be bound by Pdx1 and NeuroD1 and verified that they were bound in my cells, and also interrogated the same sites to see if they were occupied by MafA (Figure 12), with Insulin as a positive control. MafA bound to almost all the same sites as Pdx1 and NeuroD1,

with a lesser fold enrichment for all of them, save one. This showed support for the hypothesis that these three transcription factors work together to co-regulate multiple genes, not just at the Insulin promoter. Pdx1, NeuroD1 and MafA all bound strongly to the *Il1-R1* gene, which is the signal transducing receptor for IL-1 (a principal proinflammatory cytokine), and it has been identified in genetic studies to be highly associated with T1DM[210]. 10 of 19 genes are unannotated transcripts but they showed strong binding of all three transcription factors. A few of them (e.g. *M12C1*, chr4:5499) were highly conserved across vertebrates, perhaps indicative of the presence of enhancers or previously uncharacterized regulatory regions (Figure 13).

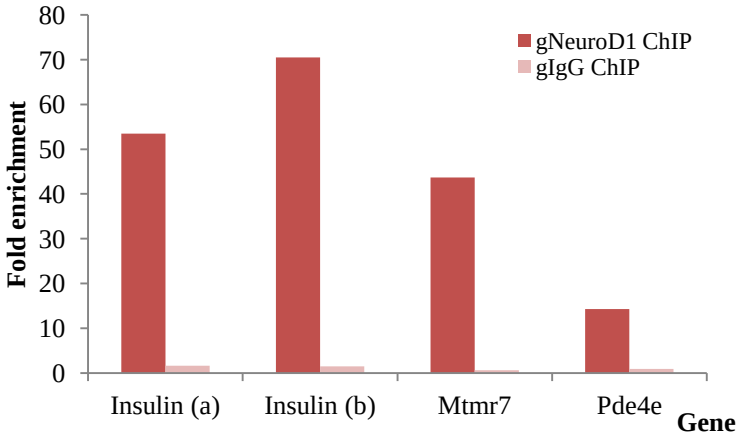
(A)



(B)



(C)



(D)

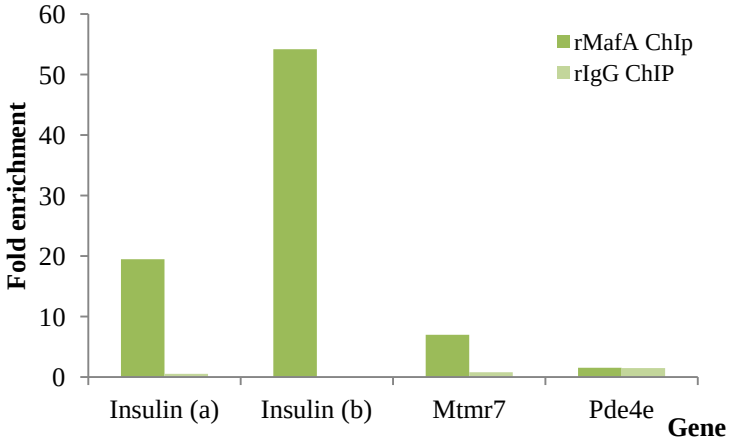


Figure 11. Specific ChIP of Pdx-1, NeuroD1 and MafA anti-serum. (A) NIT-1 chromatin sheared to a 100-500bp range of DNA fragments. ChIP from NIT-1 cells was performed using antibodies against (B) Pdx1, (C) NeuroD1 and (D) MafA, with rabbit or goat IgG antibodies as a non-specific antibody control. Fold enrichment represents the abundance of enriched DNA fragments over a control region (Utf1) shown to not be bound by Pdx1 and NeuroD1. Each ChIP experiment was performed with three technical replicates for each RT-PCR, and the average of the normalized ratio to the target gene/*Utf1* was calculated and presented. This data is representative of ChIP experiments performed in three biological replicates.

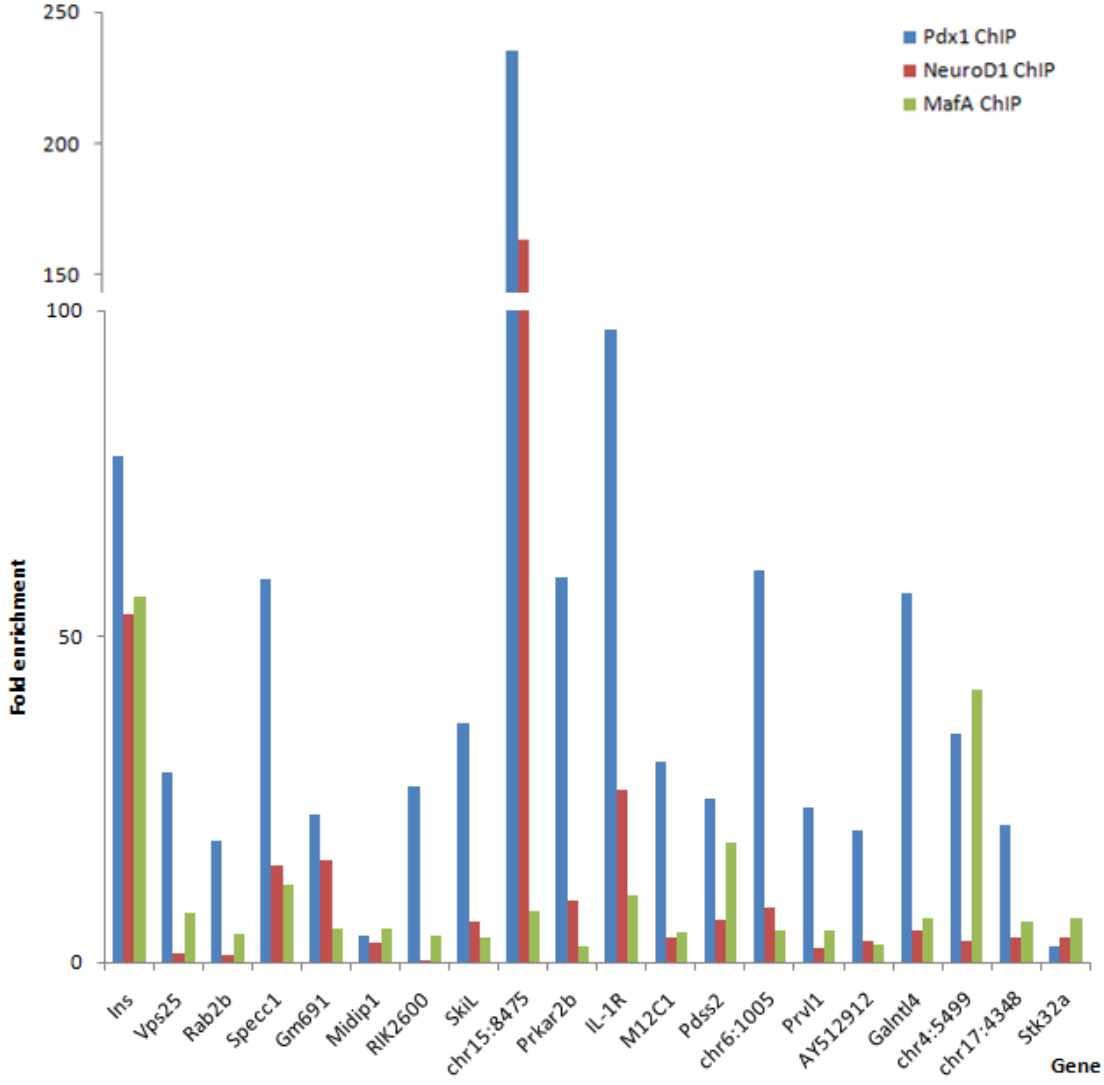
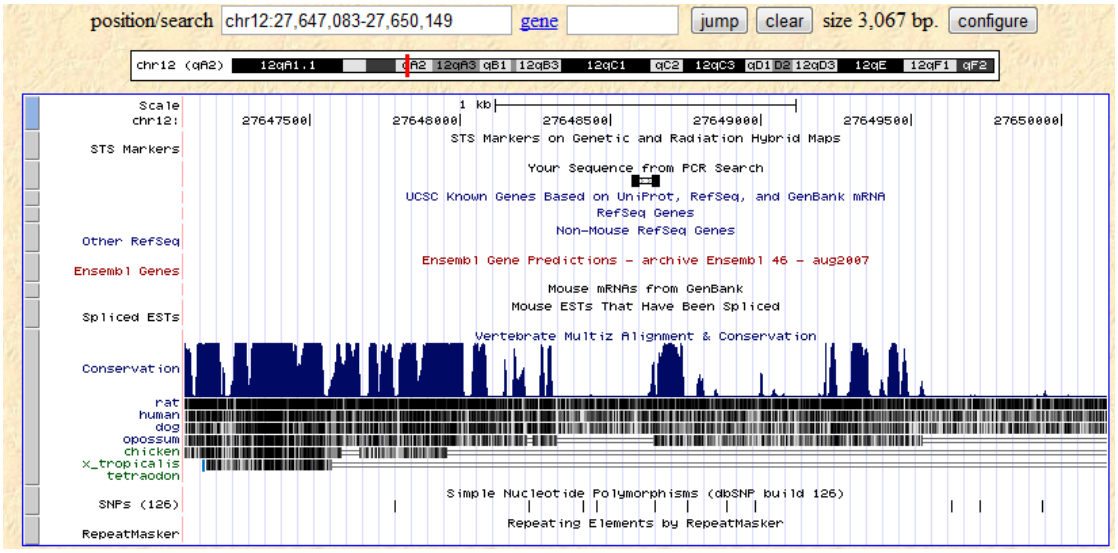


Figure 12. Pdx1, NeuroD1 and MafA binds to sites previously identified as Pdx1 and NeuroD1 targets[86]. ChIP from NIT-1 cells was performed using antibodies against Pdx1, NeuroD1 and MafA, in the same manner as Figure 11. Each ChIP experiment was performed with three technical replicates for each RT-PCR, and the average of the normalized ratio to the target gene/*Utf1* was calculated and presented.

(A) M12C1



(B) chr4:55,767,774-55,774,703

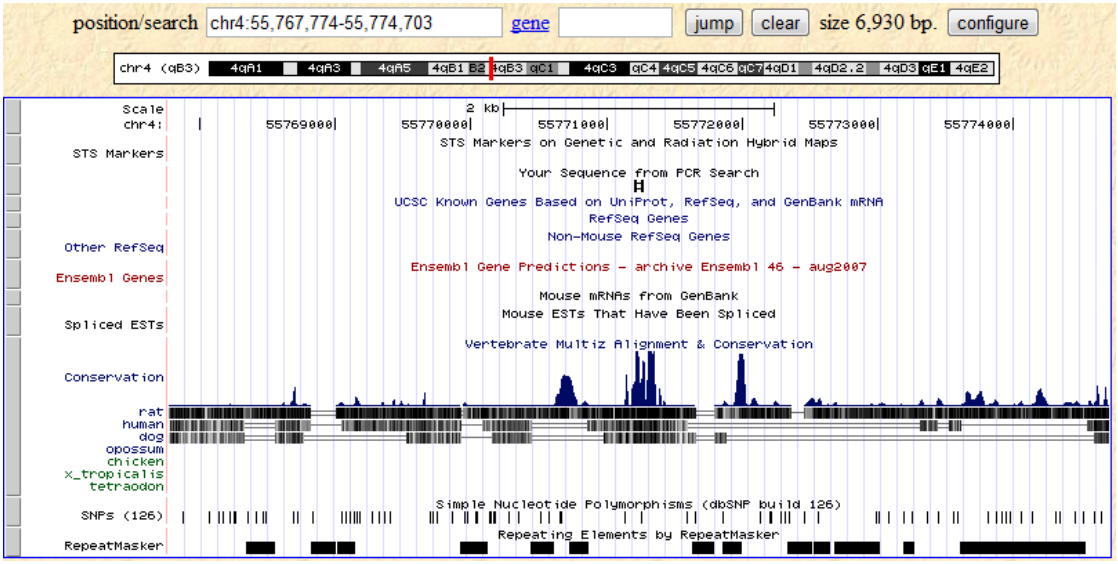


Figure 13. Regions of the mouse genome identified to be bound by Pdx1, NeuroD1 and MafA are highly conserved across species at the (A) *M12C1* and (B) chr4:55,767,774-55,774,703 locus.

3.2.2. Global Transcriptional Binding Landscape of Pdx1, NeuroD1 and MafA

I extended the search for Pdx1, NeuroD1 and MafA target genes to a genome-wide level by performing ChIP-Seq on NIT-1 cells with specific antibodies. I performed the ChIPs in biological duplicates (i.e. chromatin was harvested from two separate batches of cells and ChIPs performed on different days) and validated the quality of the ChIPs in the same fashion as Figure 11 before preparing the sample for sequencing. Input DNA was sequenced as a control for background noise, and subsequent reads mapped against that. The number of mapped reads per 10bp window (genome-wide) of each biological replicate was counted, and the correlation between each transcription factor ChIP-Seq library was calculated. Pdx1, NeuroD1 and MafA ChIP-Seq biological replicates show 0.99, 0.97 and 0.97 correlation with one another respectively (Figure 14).

As the biological replicates were highly correlated with one another, the raw reads from each biological replicate were pooled together and these combined reads were used for subsequent analyses (Table 2). HOMER was used to perform peak calling, binding peaks were defined as those above the FDR (False Discovery rate) cutoff of 0.001 and Poisson p-value ~ 2.7^{-07}. 23132 peak regions were identified to be Pdx1-bound, 31031 peak regions NeuroD1-bound and 24108 peak regions MafA-bound. The unique reads were also put through MACS and similar numbers of peaks were identified (Pdx1:26153, NeuroD1:34507, MafA:26234). Pdx1, NeuroD1 and MafA were shown to bind to one another's promoters, and the *Insulin* promoter was also bound by all three transcription factors (Figure 15). To validate some of the binding sites, 10 peaks with a significant p-value were selected from all the binding sites identified to be bound by each of the three transcription factors, and another 10 binding sites were randomly selected to be bound by each of the three transcription factors. I designed primers against these regions (20 for each transcription factor) and new ChIPs were prepared and ChIP-qPCRs were run. The positive enrichments overall (Figure 16) shows that the ChIP-Seq libraries generated were generally reliable and had identified *bona fide* binding sites.

Transcription factor MACS peaks were overlapped to the University of California, Santa Cruz (UCSC) genome data browser, using the Feburary 2006 (NCBI36/mm8) *Mus musculus* reference assembly. The peaks regions were visualized (Figure 17) and are found to be evenly distributed across the genome, with no preference for particular chromosomes. The peaks were next annotated based on the genomic region that they overlapped to: Promoter, intron, exon, intergenic, 5'UTR and 3'UTR. The relative enrichments of ChIP regions of Pdx1, NeuroD1 and MafA (Figure 18) are displayed in important genomic features, such as promoters, immediate downstream of genes, and gene bodies, with respect to the genome background. Pdx1 shows very high relative enrichment in promoters, with 15% of ChIP regions enriched at the promoter compared to 2.7% of the genome background but very low enrichment in bidirectional ones, with only 1.7% of ChIP regions found within 3 kb upstream of TSS (Figure 18A). NeuroD1 and MafA had 6.4% and 5.7% of their peaks in the promoters (Figure 18B). A different representation of the genome-wide peak annotation is displayed in Figure 19, where transcription factor peak calling with the HOMER algorithm was used. A similar result was obtained, with 17% of peaks for Pdx1 found in the promoter, compared to 6% of peaks in NeuroD1 and 6% of peaks in MafA.

The average binding sites of Pdx1, NeuroD1 and MafA were mapped around the promoter regions (±6kb from the TSS), and a peak of enrichment was observed around the promoter for all the transcription factors, indicating they bound together at the promoter (Figure 20). Pdx1, NeuroD1 and MafA peaks were over-lapped with Pdx1, NeuroD1 and MafA peak centers. There was enrichment around all the three peak centers, suggesting that they associate strongly with each other (Figure 21).

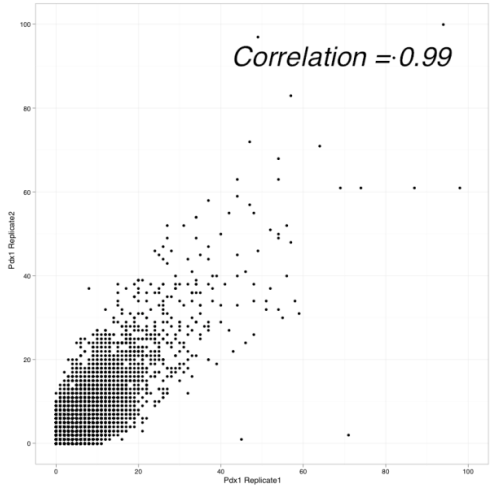
	Pdx1	NeuroD1	MafA
Replicate 1	9062661	12537461	27396468
Replicate 2	12260471	17092581	18295806
Total	21323132	29630042	27396468
HOMER	23132	31031	24108
MACS	26153	34507	26234

	Total peaks	3' UTR	Promoter	TSS	Exon	Intron	Intergenic	5' UTR
Pdx1	26153	87	4013	239	533	8598	9561	101
NeuroD1	34507	137	2009	287	676	13759	14106	57
MafA	26234	110	1378	228	567	10783	10991	51

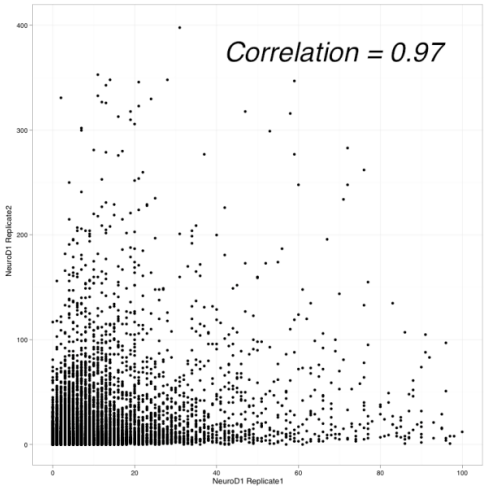
Table 2. Mapping and peak calling of Pdx1, NeuroD1 and MafA ChIP-Seq libraries.

Pdx1, NeuroD1 and MafA ChIP-Seq was performed in NIT-1 cells and the reads were mapped to the mm8 genome with ELAND. Using HOMER, there were 23132 Pdx1 peaks, 34507 NeuroD1 peaks and 26234 MafA peaks. (Gireesh Bogu performed the analysis.)

(A) Pdx1



(B) NeuroD1



(C) MafA

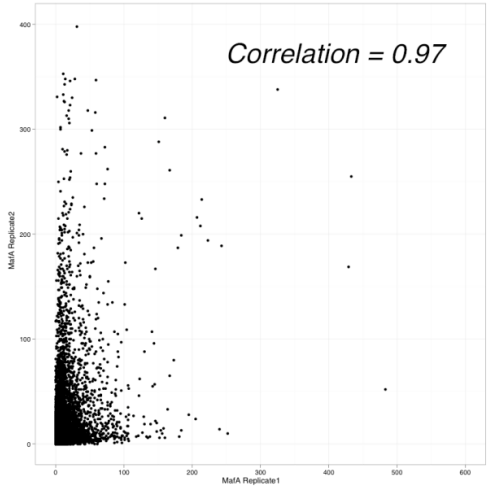
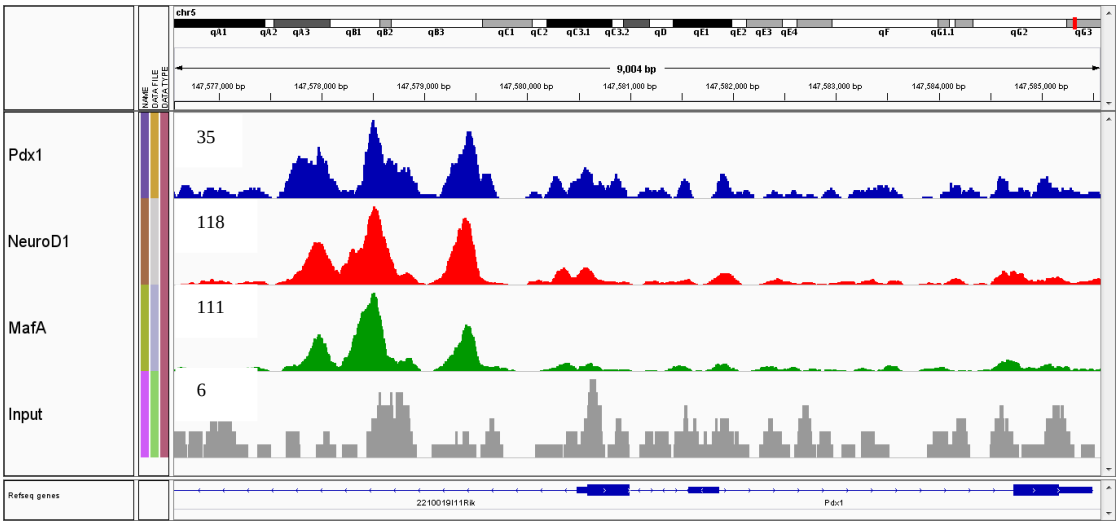
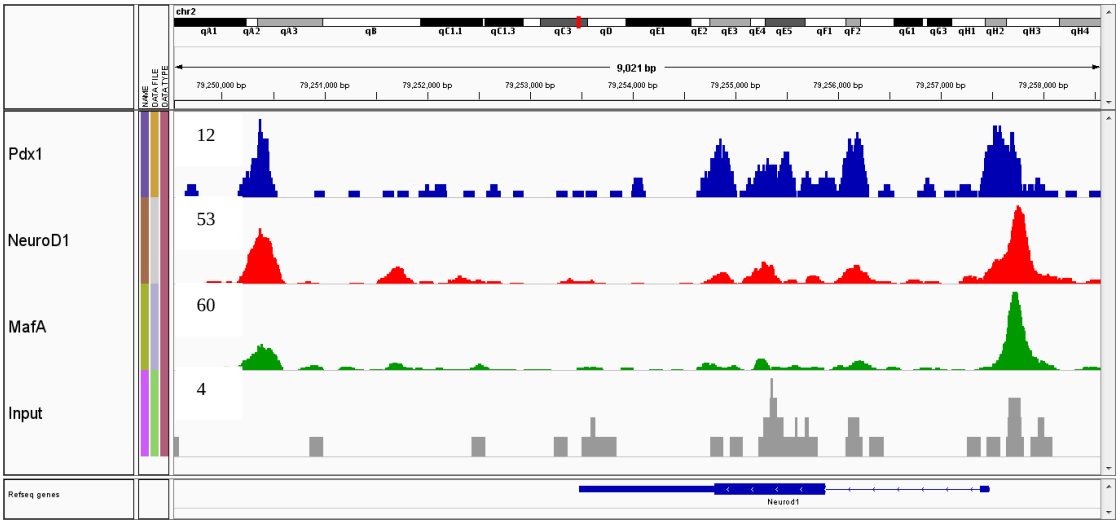


Figure 14. Biological replicates of Pdx1, NeuroD1 and MafA ChIP-Seq libraries are highly correlated with one another. The number of mapped reads per 10bp window (genome-wide) were counted, and the correlation between replicates was calculated. (A) Pdx1, (B) NeuroD1 and (C) MafA ChIP-Seq biological replicates show 0.99, 0.97 and 0.97 correlation with one another respectively. (Gireesh Bogu performed the analysis.)

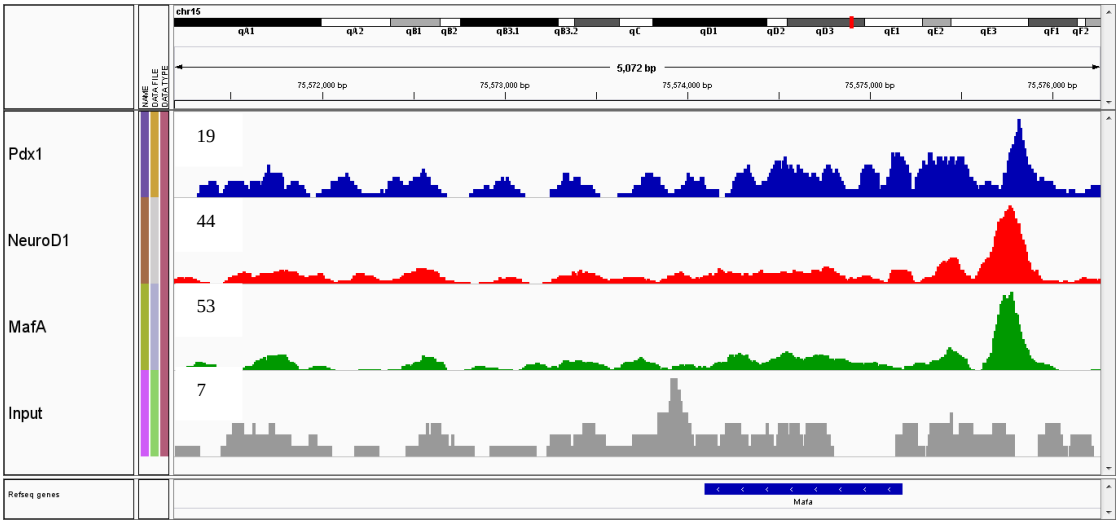
(A) Pdx1



(B) NeruoD1



(C) MafA



(D) Ins1

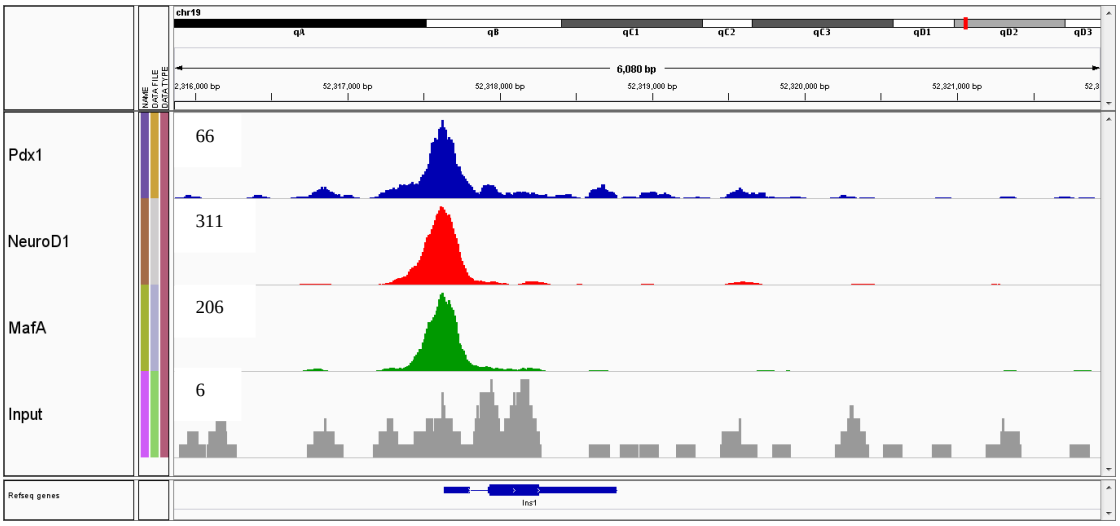
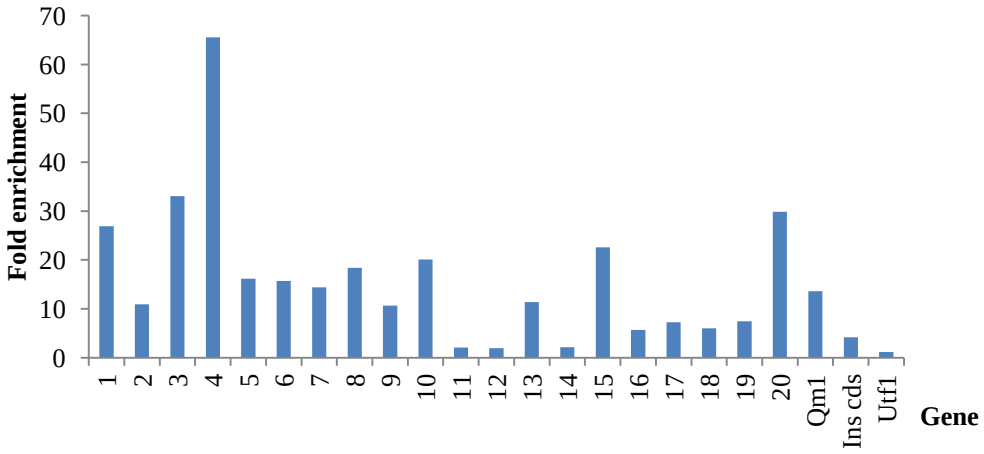
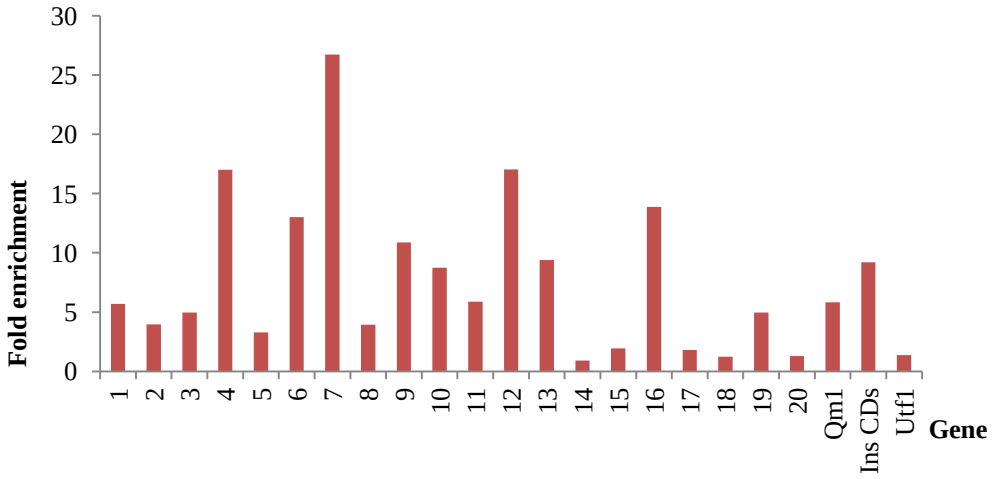


Figure 15. Pdx1, NeuroD1 and MafA binding at the *Pdx1*, *NeuroD1*, *MafA* and *Ins1* gene. Pdx1, NeuroD1 and MafA ChIP-Seq was performed in NIT-1 cells and the binding of all three transcription factors at the (A) *Pdx1*, (B) *NeuroD1*, (C) *MafA* and (D) *Insulin1* promoter was visualized with the Integrated Genomics Viewer (IVG)[211] software. The input library was included as a control for background signal. Numbers indicate the peak height.

(A) Pdx1



(B) NeuroD1



(C) MafA

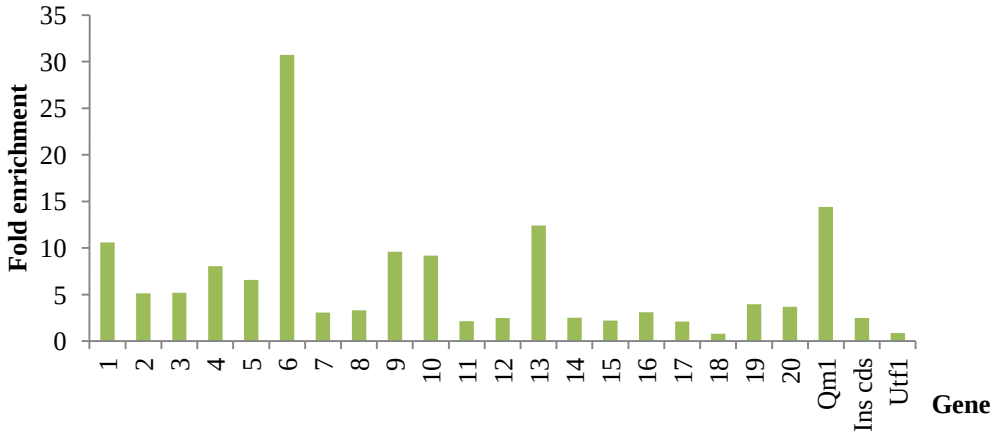
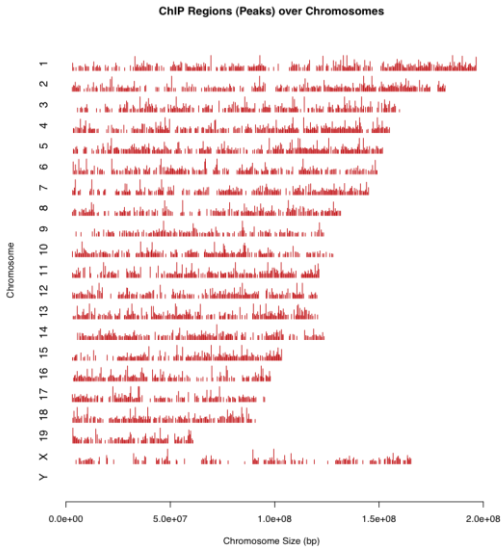
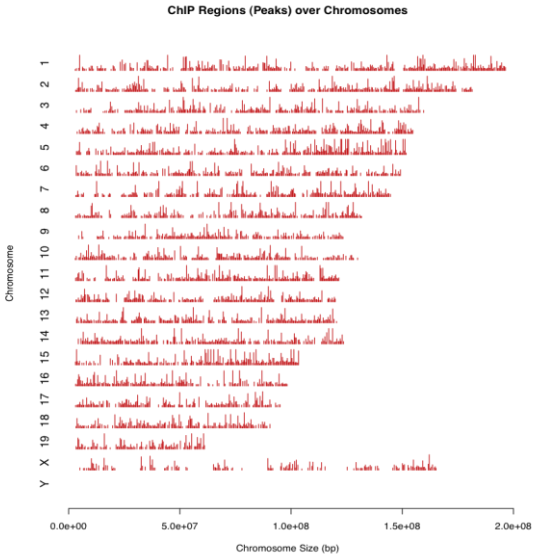


Figure 16. Confirmation of Pdx1, NeuroD1 and MafA ChIP-Seq bound genes by ChIP-qPCR. ChIP from NIT-1 cells was performed using antibodies against (A) Pdx1, (B) NeuroD1 and (C) MafA, in the same manner as Figure 11. Genes 1-10 have a highly significant p-value and genes 11-20 were randomly picked from all the regions identified to be bound by each transcription factor. Each ChIP experiment was performed with three technical replicates for each RT-PCR, and the average of the normalized ratio to the target gene/*Utf1* was calculated and presented.

(A) Pdx1



(B) NeuroD1



(C) MafA

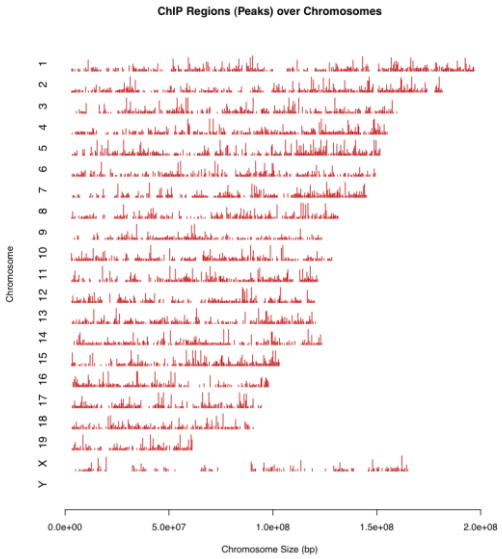
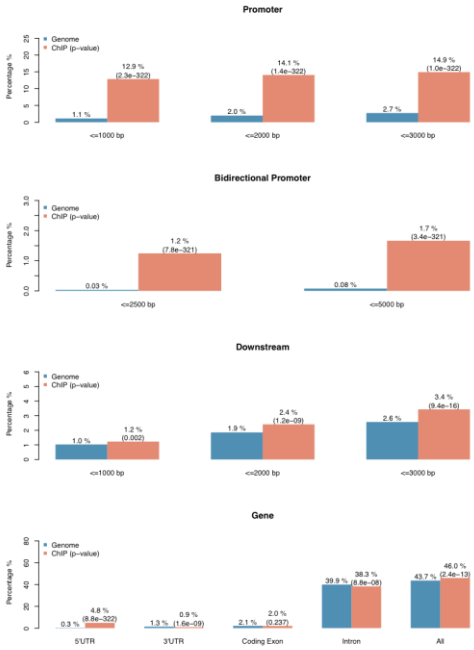
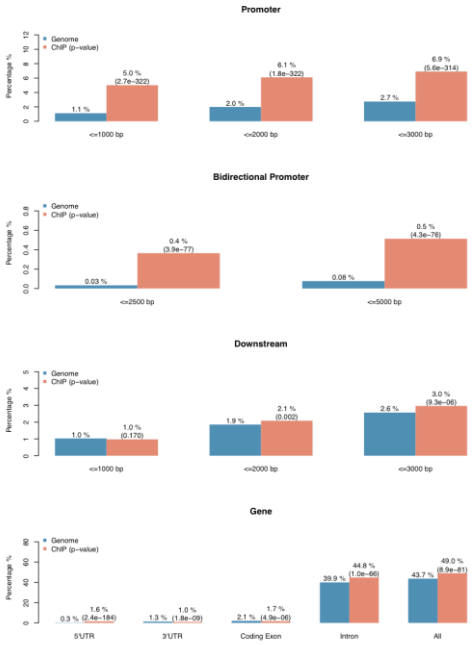


Figure 17. Distribution of ChIP regions of Pdx1, NeuroD1 and MafA over the genome along with their scores or peak heights. The line graph illustrates the distribution of (A) Pdx1, (B) NeuroD1 and (C) MafA peak heights (or scores). The red bars in the main plot represent ChIP regions. The x-axis of the main plot represents the actual chromosome sizes (in base pairs) and the y-axis represents chromosome names. (Gireesh Bogu performed the analysis.)

(A) Pdx1



(B) NeuroD1



(B) MafA

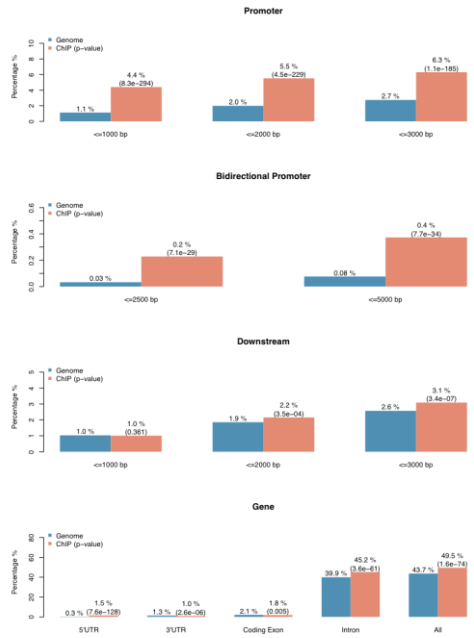


Figure 18. Relative enrichments of ChIP regions of Pdx1, NeuroD1 and MafA in important genomic features, such as promoters, immediate downstream of genes, and gene bodies, with respect to the genome background. The blue bars represent the percentages of the tiled or mappable regions that are located in such genomic regions (genome background) and the red bars the percentages of ChIP regions. (A) Pdx1 shows very high relative enrichment in promoters (15% of ChIP regions compared to 2.7 % of the genome background) but very low enrichment in bidirectional ones (only 1.7 % of ChIP regions within 3 kb up stream of TSS). (B) NeuroD1 shows 6.1% enrichment in promoters. (C) MafA has 5.5% of its peaks enriched in promoters. (Gireesh Bogu performed the analysis.)

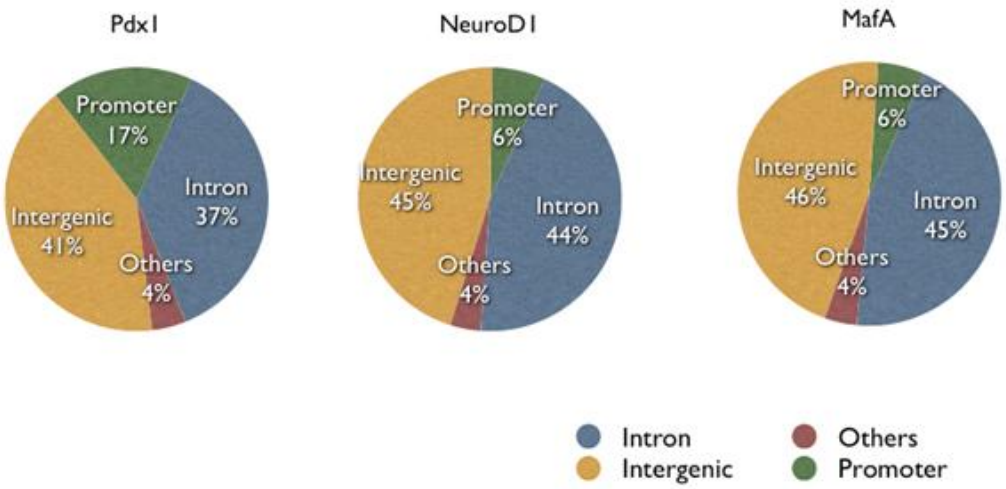


Figure 19. Relative enrichments of ChIP regions of Pdx1, NeuroD1 and MafA in important genomic features. ChIP-Seq peaks were uploaded to UCSC mm8 refGene and annotated based on the genomic location (promoter, intron, exon, intergenic, 5'UTR and 3'UTR) they overlapped to. 17% of peaks in Pdx1 are localized to the promoter in contrast to 6% in NeuroD1 and MafA. (Gireesh Bogu performed the analysis.)

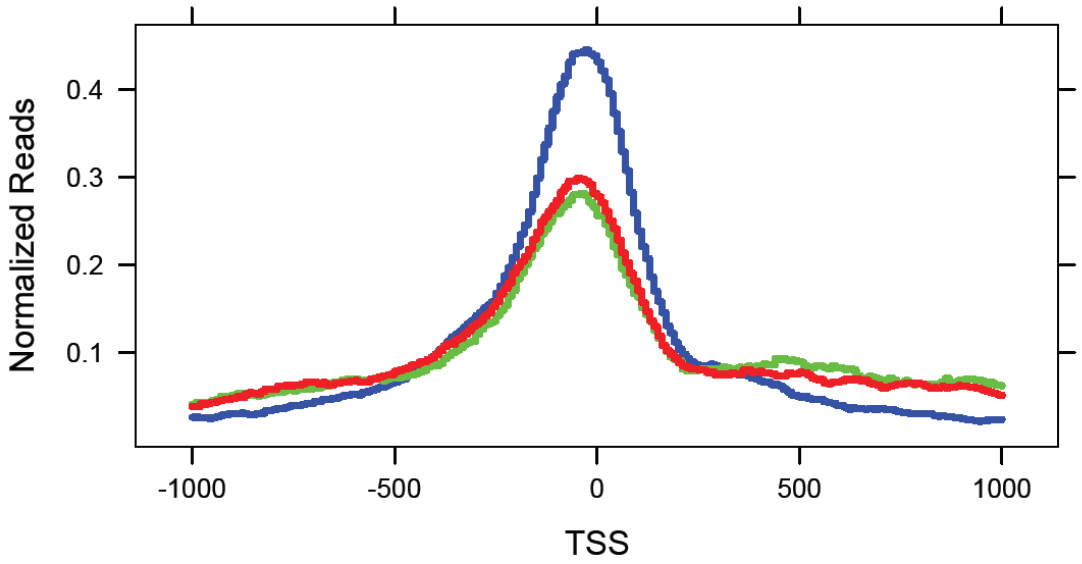
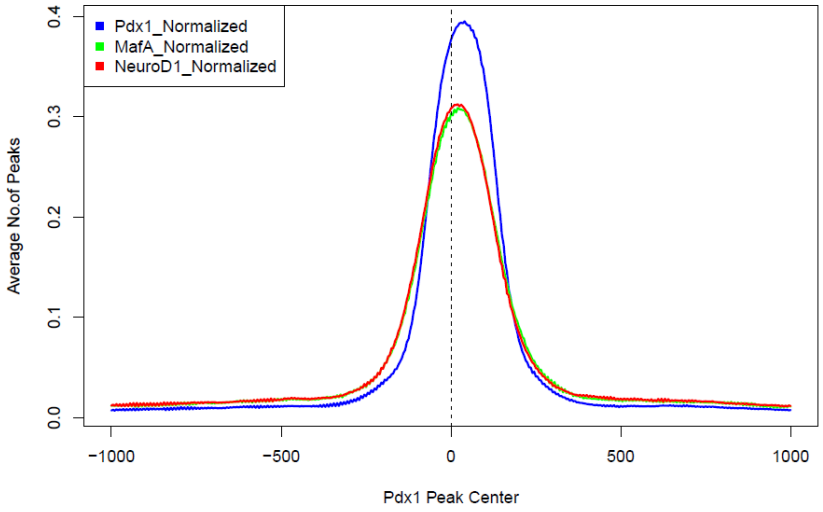
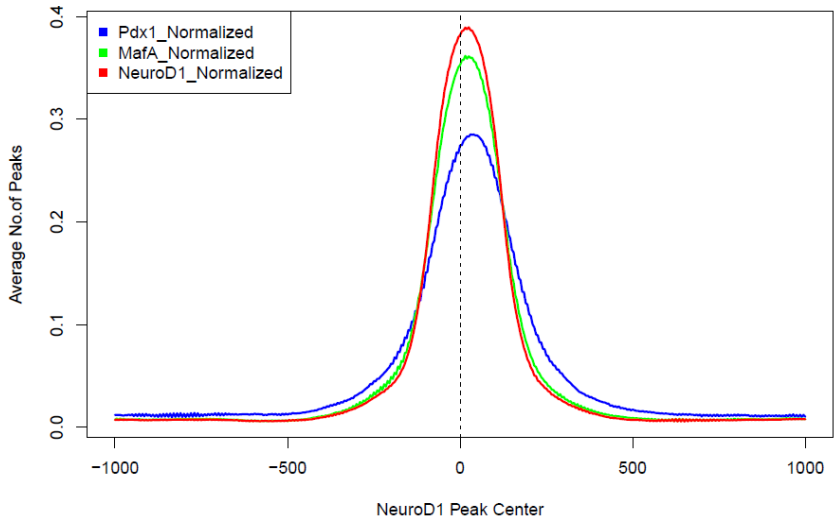


Figure 20. Pdx1, NeuroD1 and MafA are enriched around the TSS. Average binding sites of Pdx1 (blue), Neurod1 (red) and MafA (green) were mapped around the promoters (±6 kb) and a peak for all three transcription factors was observed at the promoter, indicating they are bound there. (Gireesh Bogu performed the analysis.)

(A) Pdx1



(B) NeuroD1



(C) MafA

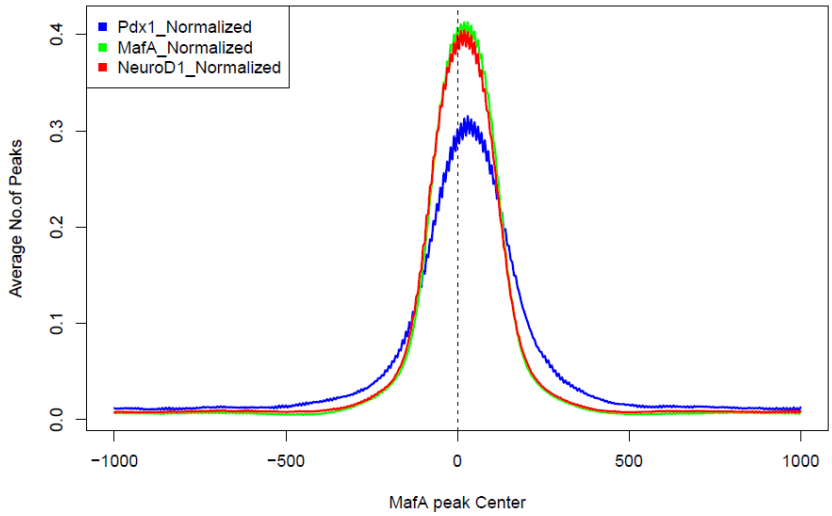


Figure 21. Spatial distribution of the distance (in base pairs) of Pdx1, NeuroD1 and MafA. Pdx1 peaks, NeuroD1 and MafA peaks are overlapped with (A) Pdx1, (B) NeuroD1 and (C) MafA peak centers. The Pdx1, NeuroD1, MafA enrichment around all 3 peak centers suggest they associate with strongly with each other. (Gireesh Bogu performed the analysis.)

To identify the DNA binding sequences associated with each transcription factor, a *de novo* motif discovery strategy was employed. ChIP-Seq peaks of Pdx1, NeuroD1 and MafA were input into HOMER, which extracted peak sequences with less than 70% repeats and accounted for GC biases by using randomly selected background sequences matching GC characteristics of input sequences. *De novo* motifs for Pdx1, NeuroD1 and MafA were thus identified (Figure 22). The top motif for Pdx1 had a p value of 1E-523, and the motifs for NeuroD1 and MafA were almost identical, with p values of 1E-3395 and 1E-3104 respectively. The motifs were also checked against the TRANSFEC database to see if they matched known motifs. The previously identified Pdx1 motif was among the top three motifs, with a p value of 1E-260. The motifs for NeuroD1 and MafA have not been described, and the motif for the transcription factor Atoh1 came up on top (p value 1E-3395). Interestingly, the motif for the general transcription factor E2A was among the top five motifs for both NeuroD1 and MafA, with p values of 1E-554 and 1E-547 respectively. A full list of identified motifs is available in Supplementary Table 1.

I next wanted to ask if these transcription factors bound alone or together across the genome. To this end, Pdx1, NeuroD1 and MafA ChIP-Seq peaks were resized to ± 65 bp from the peak center and overlapped by a single base pair. There were 12780 Pdx1-only peaks, 8828 NeuroD1-only and 3473 MafA-only peaks (Figure 23A). There were 657 peaks common to both Pdx1 and MafA, 2225 to Pdx1 and NeuroD1, and very strikingly, 12508 peaks were bound by both MafA and NeuroD1. There were 7470 common peak regions bound by all three transcription factors. The same analysis was also performed at the promoters of genes (±2kb from transcription start site). Figure 23B shows that the same trend was seen, with a larger number of Pdx1 only peaks than NeuroD1 or MafA only (3094 vs 682 and 299 respectively). There was also the same striking number of peaks co-occupied by both NeuroD1 and MafA (798), though the number of peaks co-occupied by Pdx1 and NeuroD1 are now similar (843). There were 1306 peaks with all three transcription factors binding.

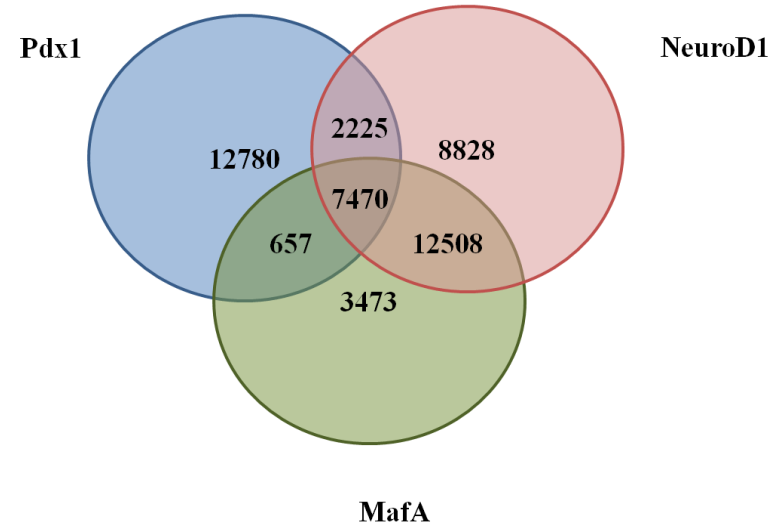
(A) Pdx1

(B) NeuroD1

(C) MafA

Figure 22. DNA binding motifs discovered for Pdx1, NeuroD1 and MafA. Fixed-width peaks from (A) Pdx1, (B) NeuroD1 and (C) MafA ChIP-Seq peaks (±50bp from peak center) were input into HOMER for de novo motif discovery. The Pdx1 top motif had a p-value $1e^{-523}$, the top motif of NeuroD1 had a p-value $1e^{-3395}$, and the top motif for Mafa had a p-value of $1e^{-3104}$. (Gireesh Bogu performed the analysis.)

(A)



(B)

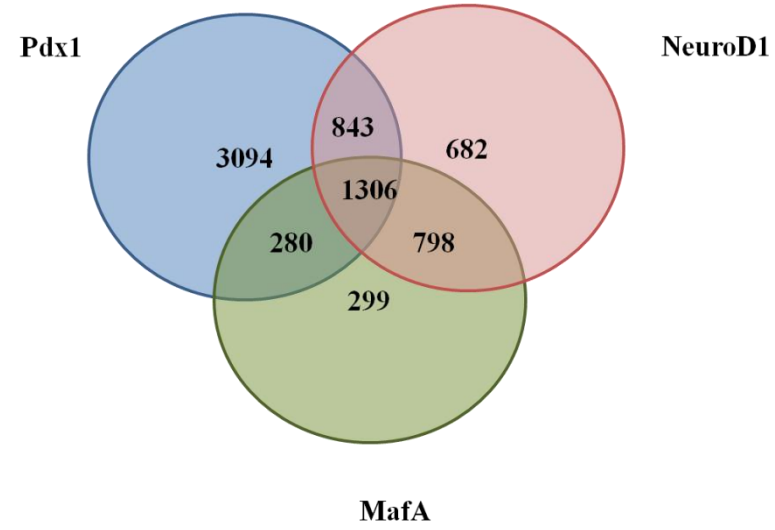


Figure 23. Co-occupancy of Pdx1, NeuroD1 and MafA. Genomic regions bound by Pdx1, NeuroD1 and MafA either individually or together were identified by resizing peaks to ±65bp from peak center, with an overlap by a single base pair. (A) Co-occupancy of Pdx1, NeuroD1 and MafA binding sites with all ChIP-Seq peaks, and (B) shows co-occupancy at promoters of genes (±2kb from transcription start site). Pdx1 has more independent sites, while there is a big overlap of NeuroD1 and MafA binding sites. (Gireesh Bogu performed the analysis.)

3.2.3. PDX1, NEUROD1 & MAFA ARE BOUND TO GENES INVOLVED IN SIMILAR BIOLOGICAL PROCESSES

To shed light onto biological relevance of the network of genes bound by Pdx1, NeuroD1 and MafA, I uploaded the RefSeq list of nearest bound genes by each transcription factor into Ingenuity Pathway Analysis (IPA, Ingenuity® Systems)[187] and identified the top biological networks and pathways associated with the transcription factor binding sites for Pdx1 (Table 3), NeuroD1 (Table 4) and MafA (Table 5). Each identifier was mapped to its corresponding object in the Ingenuity® Knowledge Base and termed network eligible molecules. These were then overlaid onto a global molecular network developed from information contained in the Ingenuity Knowledge Base, and networks were then algorithmically generated based on their connectivity. The functional analysis identified the biological functions and/or diseases that were most significantly enriched in the data set.

Top disease networks associated with Pdx1 binding sites include those of genetic disorders ($p<4.04^{-75}$), neurological ($p<1.85^{-57}$), metabolic ($p<6.39^{-46}$), gastrointestinal ($p<2.39^{-43}$) and endocrine diseases ($p<1.75^{-39}$) (Table 3). Pdx1 binding sites are also highly enriched in cellular regulatory activities, e.g. cell death ($p < 2.65^{-32}$), cell cycle ($p<4.31^{-29}$), cellular development ($p<9.39^{-18}$), and it is involved in embryonic ($p<9.2^{-12}$) and organ development ($p<2.05^{-13}$), nervous system development and function ($p<1.45^{-12}$). A full list of genes involved in those pathways is listed in Supplementary Table 2A.

NeuroD1-bound genes are associated with lipid and amino acid metabolism, endocrine system development and function, and general cellular functions (Table 4). Disease networks associated with NeuroD1 binding sites include genetic disorders ($p<3.87^{-104}$), neurological diseases ($p<2.81^{-77}$), gastrointestinal disease ($p<3.5^{-64}$), and metabolic disease ($p<7.06^{-56}$). Like Pdx1, NeuroD1-bound genes appear to be involved in general cellular regulatory activities, e.g. cell death ($p<5.32^{-31}$), cellular growth and proliferation ($p<1.45^{-26}$) and cellular development ($p<5.17^{-16}$). Genes bound by NeuroD1 also appear to be involved in organ

development ($p<1.66^{-16}$), nervous system and development ($p<9.08^{-16}$) and tissue development ($p<4.01^{-15}$). A full list of genes involved in those pathways is listed in Supplementary Table 2B.

Genes bound by MafA are associated with infectious diseases, lipid metabolism, and general cellular processes and maintenance (Table 5). MafA disease-related networks are similar to that of Pdx1 and NeuroD1, comprising of genetic disorders ($p<2.22^{-107}$), neurological diseases ($p<7.43^{-83}$), gastrointestinal disease ($p<8.19^{-67}$), and metabolic disease ($p<4.96^{-62}$). Similarly, MafA-bound genes also appeared to be involved in general cellular regulatory activities, e.g. cell death ($p<4.18^{-33}$), cellular growth and proliferation ($p<2.47^{-25}$) and cellular development ($p<7.43^{-21}$). Genes bound by MafA also are linked to in organ development ($p<5.63^{-19}$), nervous system and development ($p<4.39^{-17}$) and organismal survival ($p<4.51^{-16}$). A full list of genes involved in those pathways is listed in Supplementary Table 2C.

The similarities in the biological networks and pathways identified suggests that Pdx1, NeuroD1 and MafA share common regulatory roles in maintaining the β-cell phenotype, by means of organ and nervous system development and cell cycle activities.

Associated Network Functions	Network Score
Gene Expression, Cellular Assembly and Organization, Cellular Compromise	20
RNA Post-Transcriptional Modification, Protein Synthesis, Cellular Assembly and Organization	20
Dermatological Diseases and Conditions, Genetic Disorder, Hair and Skin Development and Function	20
RNA Post-Transcriptional Modification, Gene Expression, Protein Synthesis	20
Developmental Disorder, Genetic Disorder, Neurological Disease	20

Diseases and Disorders	p-value	# Molecules
Genetic Disorder	4.04E-75	4095
Neurological Disease	1.85E-57	2466
Metabolic Disease	6.39E-46	1766
Gastrointestinal Disease	2.39E-43	2450
Endocrine System Disorders	1.75E-39	1559

Molecular and Cellular Functions	p-value	# Molecules
Cell Death	2.65E-32	1812
Gene Expression	5.76E-30	1313
Cell Cycle	4.31E-29	897
Cellular Growth and Proliferation	2.65E-20	1826
Cellular Development	9.39E-18	1576

Physiological System Development and Function	p-value	# Molecules
Organismal Survival	1.17E-16	596
Organismal Development	3.80E-15	746
Organ Development	2.05E-13	617
Nervous System Development and Function	1.45E-12	1028
Embryonic Development	9.20E-12	473

Table 3. Top biological networks and pathways associated with Pdx1 target genes.

Pdx1 ChIP-Seq peaks were identified with MACS and IPA was used to generate this list of networks associated with its biological, molecular and cellular function, physiological system development and function. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible molecules in the dataset.

Associated Network Functions	Network Score
Lipid Metabolism, Small Molecule Biochemistry, Vitamin and Mineral Metabolism	20
Amino Acid Metabolism, Post-Translational Modification, Small Molecule Biochemistry	20
Endocrine System Development and Function, Tissue Morphology, Cellular Growth and Proliferation	20
Cellular Assembly and Organization, Cell Cycle, DNA Replication, Recombination, and Repair	20
Gene Expression, Cardiovascular System Development and Function, Organ Development	20

Diseases and Disorders	p-value	# Molecules
Genetic Disorder	3.87E-104	4267
Neurological Disease	2.81E-77	2580
Gastrointestinal Disease	3.50E-64	2455
Cardiovascular Disease	1.59E-60	1656
Metabolic Disease	7.06E-56	1842

Molecular and Cellular Functions	p-value	# Molecules
Cell Death	5.32E-31	1845
Cellular Growth and Proliferation	1.45E-26	1901
Gene Expression	8.58E-26	1264
Cell Cycle	1.16E-18	866
Cellular Development	5.17E-16	1638

Physiological System Development and Function	p-value	# Molecules
Organ Development	1.66E-16	653
Nervous System Development and Function	9.08E-16	1116
Tissue Development	4.02E-15	1025
Organismal Development	1.95E-14	969
Organismal Survival	7.18E-14	594

Table 4. Top biological networks and pathways associated with NeuroD1 target genes.

NeuroD1 ChIP-Seq peaks were identified with MACS and IPA was used to generate this list of networks associated with its biological, molecular and cellular function, physiological system development and function. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible molecules in the dataset.

Associated Network Function	Network Score
Infection Mechanism, Infectious Disease, Antigen Presentation	22
RNA Post-Transcriptional Modification, Skeletal and Muscular System Development and Function, Tissue Development	22
Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization, Cellular Function and Maintenance	22
Lipid Metabolism, Small Molecule Biochemistry, Cellular Development	22
Cell Morphology, Cellular Function and Maintenance, DNA Replication, Recombination, and Repair	22

Diseases and Disorders	p-value	# Molecules
Genetic Disorder	2.22E-107	3841
Neurological Disease	7.43E-83	2350
Gastrointestinal Disease	8.19E-67	2224
Cardiovascular Disease	2.24E-65	1522
Metabolic Disease	4.96E-62	1689

Molecular and Cellular Functions	p-value	# Molecules
Cell Death	4.18E-33	1662
Cellular Growth and Proliferation	2.47E-25	1727
Gene Expression	2.12E-24	1129
Cell Cycle	5.60E-21	792
Cellular Development	7.43E-21	1498

Physiological System Development and Function	p-value	# Molecules
Tissue Development	2.17E-20	900
Organ Development	5.63E-19	615
Organismal Development	2.30E-18	806
Nervous System Development and Function	4.39E-17	1035
Organismal Survival	4.51E-16	549

Table 5. Top biological networks and pathways associated with MafA target genes.

MafA ChIP-Seq peaks were identified with MACS and IPA was used to generate this list of networks associated with its biological, molecular and cellular function, physiological system development and function. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible molecules in the dataset.

3.2.4. PDX1, NEUROD1 AND MAFA REGULATE A COMMON SET OF GENES

I hypothesized that, as these transcription factors had commonalities in the function of the genes they bound to, similarities would also exist in the genes regulated by them. To test this hypothesis, RNA interference (RNAi) was performed for Pdx1, NeuroD1 and MafA in biological quadruplicate to identify genes regulated upon each transcription factor knockdown. After 48 hours of RNAi, the expression of *Pdx1*, *NeuroD1* and *MafA* was decreased to 30%, 25%, and 5% respectively (Figure 24A). Corresponding decreases in the protein levels were also observed (Figure 24B).

In order to assess global gene expression changes in the RNAi samples, Illumina bead chip arrays were used. In response to the depletion of endogenous Pdx1, 570 genes were significantly regulated, of which 403 were up- and 167 down-regulated (±1.5 fold, $p<0.05$). In response to the depletion of NeuroD1, 1550 genes were significantly regulated, of which 946 were up- and 604 down-regulated (±1.5 fold, $p<0.05$). When MafA protein levels were reduced, 552 genes were significantly regulated, of which 341 were up- and 212 down-regulated (±1.5 fold, $p<0.05$). A complete list of genes in each of the microarray datasets is available in Supplementary Table 3.

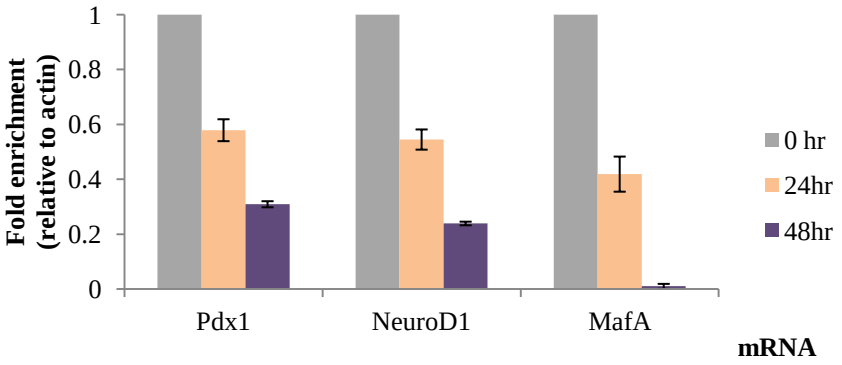
To look at how Pdx1, NeuroD1 and MafA genes were co-regulated, genes that changed upon each transcription factor knockdown were overlapped (Figure 25) and compared. Strikingly, 834 genes were regulated by NeuroD1 only, in comparison to the 175 and 151 regulated by Pdx1 and MafA respectively. There was also a lesser number of genes found to be regulated by both Pdx1 and MafA (39), when compared to Pdx1 and NeuroD1 (84) and MafA and NeuroD1 (86).

The comparatively large number of NeuroD1-only genes suggested that NeuroD1 had more indirect targets, which led me to ask whether the target genes of NeuroD1 consisted of more

transcription factors, when compared to Pdx1 and MafA regulated genes. To address this, the genes identified to be regulated by Pdx1, NeuroD1 and MafA in the RNAi-depleted genes were overlapped with a list of 1700 known transcription factors (courtesy of Dr Shaym Pyrakar, Genome Institute of Singapore). 12.5% of NeuroD1 target genes were identified as transcription factors, compared to 8.5% in Pdx1 and MafA target genes (Table 6).

To gain insight into the biological processes regulated by Pdx1, NeuroD1 and MafA, the list of genes identified in the microarrays was uploaded into the PANTHER database[212]. In this analysis, the genes were clustered into functional themes and compared to the *Mus musculus* reference list to look for statistically over (+) and under (-) represented pathways. Out of all the pathways enriched, there were three broad categories: metabolic processes, developmental processes and cellular processes. Supplementary Table 4 provides the full lists of genes that are involved in each pathway. These annonations suggest that Pdx1, NeuroD1 and MafA interact in a similar fashion to co-regulate pathways involved in metabolic processes (e.g. protein metabolism, nucleoside metabolism, lipid metabolism), developmental processes (e.g. response to pheromones, system development), and cellular processes (e.g. signal transduction, vesicle-mediated transport, and cell cycle).

(A)



(B)

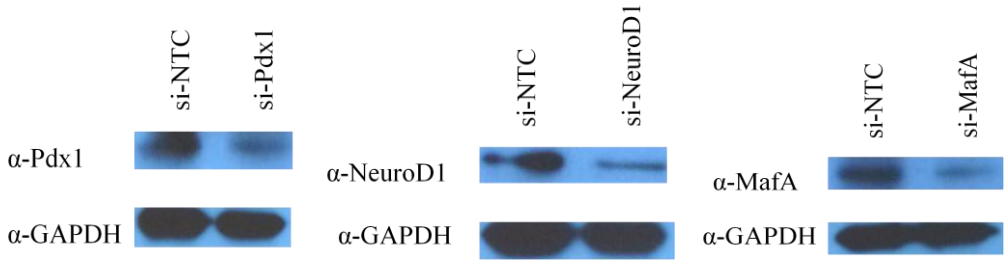


Figure 24. siRNA depletion of endogenous *Pdx1*, *NeuroD1* and *MafA* mRNA and protein levels. (A) *Pdx1*, *NeuroD1* and *MafA* levels were depleted by RNA interference using siRNA against each transcription factor in NIT-1 cells. RNA was harvested between 24hrs and 48hrs post-transfection and transcript levels assayed by RT-PCR. Mean values ± S.E. are plotted as percentages relative to the non-targeting control (NTC) (100%). The siRNA experiments were performed in quadruplicate and normalized to endogenous *β-actin*. (B) Corresponding decrease in protein levels following Pdx1, NeuroD1 and MafA RNAi treatment. Gapdh protein was used as a loading control.

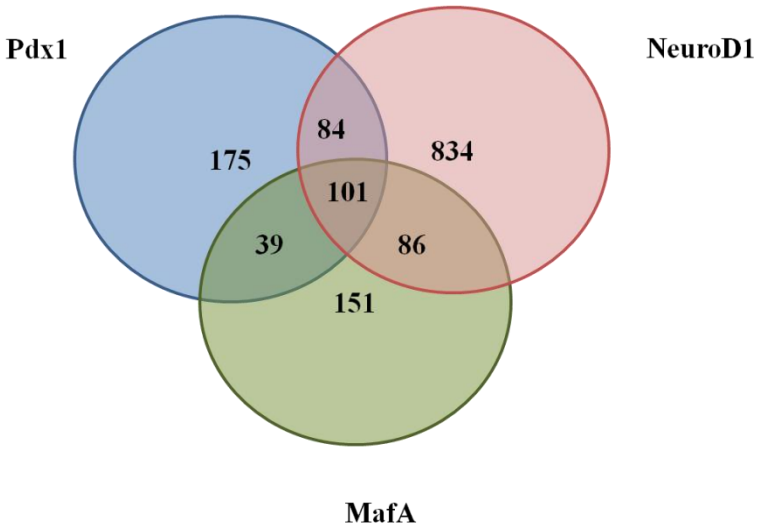


Figure 25. Significant overlap of Pdx1, NeuroD1 and MafA regulated genes. Pdx1, NeuroD1 and MafA RNAi genes regulated ±1.5 fold show statistically significant overlap. NeuroD1-only regulated genes were much more than Pdx1-only and MafA-only regulated genes. (Gireesh Bogu performed the analysis.)

Transcription factor	% of overlap with known transcription factors
Pdx1	8.5%
NeuroD1	12.5%
MafA	8.5%

Table 6. NeuroD1 targets more transcription factors than Pdx1 and MafA.

The genes identified to be regulated by Pdx1, NeuroD1 and MafA in the RNAi-depleted genes were overlapped with a list containing 1700 known transcription factors (Dr Shaym Pyrakar). 12.5% of NeuroD1 target genes were identified as transcription factors. (Gireesh Bogu performed the analysis.)

	Biological Process	*Mus musculus* REFLIST (26185)	Pdx1				NeuroD1				MafA			
				Expected	+/-	P value		Expected	+/-	P value		Expected	+/-	P value
Metabolic processes	Primary metabolic process	9122	289	192.3	+	2.82E-17	393	252.91	+	1.93E-26	59	38.67	+	5.84E-05
	Metabolic process	9603	296	202.44	+	4.19E-16	406	266.25	+	5.03E-26	59	40.71	+	2.93E-04
	Protein metabolic process	3979	133	83.88	+	3.13E-08	178	110.32	+	4.08E-11	29	16.87	+	2.00E-03
	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4156	136	87.61	+	7.32E-08	187	115.23	+	6.35E-12	27	17.62	+	1.37E-02
	Lipid metabolic process	1266	50	26.69	+	2.13E-05	57	35.1	+	2.92E-04	5	5.37	-	5.50E-01
Developmental processes	Developmental process	3296	107	69.48	+	3.92E-06	126	91.38	+	1.29E-04	19	13.97	+	1.01E-01
	Mitosis	662	33	13.96	+	7.18E-06	38	18.35	+	3.04E-05	7	2.81	+	2.30E-02
	System development	2222	76	46.84	+	2.34E-05	88	61.61	+	5.21E-04	9	9.42	-	5.30E-01
	Response to pheromone	440	0	9.28	-	8.66E-05	0	12.2	-	4.54E-06	0	1.87	-	1.52E-01
Cellular processes	Cellular process	7133	228	150.37	+	7.37E-13	303	197.77	+	2.92E-17	46	30.24	+	8.59E-04
	Intracellular signaling cascade	1720	79	36.26	+	9.28E-11	86	47.69	+	1.28E-07	9	7.29	+	3.06E-01
	Transport	3009	111	63.43	+	3.77E-09	145	83.43	+	3.07E-11	20	12.76	+	2.77E-02
	Cell communication	5033	162	106.1	+	6.96E-09	207	139.54	+	9.83E-10	27	21.34	+	1.09E-01
	Signal transduction	4858	155	102.41	+	3.14E-08	198	134.69	+	5.68E-09	26	20.59	+	1.17E-01
	Cell cycle	2018	80	42.54	+	4.78E-08	104	55.95	+	1.04E-09	17	8.55	+	4.98E-03
	Protein transport	1687	70	35.56	+	6.84E-08	81	46.77	+	1.48E-06	13	7.15	+	2.66E-02
	Intracellular protein transport	1687	70	35.56	+	6.84E-08	81	46.77	+	1.48E-06	13	7.15	+	2.66E-02
	Vesicle-mediated transport	1202	46	25.34	+	9.48E-05	55	33.33	+	2.54E-04	9	5.1	+	6.98E-02

Table 7. PANTHER Biological process of significantly up/down-regulated genes upon knockdown of Pdx1, NeuroD1 and MafA.

Reflist - No. of genes from Reference list that clustered under specific Panther category.

Observed - No. of transcription factor target genes that clustered under specific category

Expected - No. of transcription factor target genes that were expected under specific category

Over/Under - No. of observed genes vs. expected (indication of over or underrepresentation)

P-value - Probability that the number of genes observed in this category occurred by chance

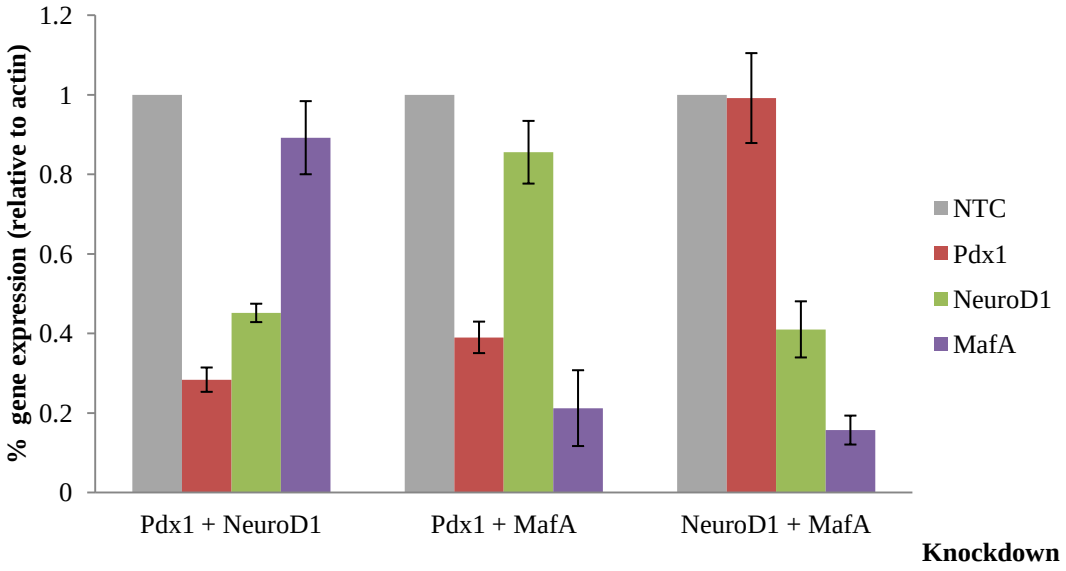
3.2.5. Combinatorial Regulation of Pdx1, NeuroD1 and MafA

As the binding sites of the three transcription factors and their downstream target genes could be categorized into similar regulatory networks, and Pdx1, NeuroD1 and MafA are found to synergistically bind to and regulate the *Insulin* promoter[124], I hypothesized that these three transcriptional factors would combinatorially regulate other genes in the genome. As shown in Figure 23, Pdx1 and MafA had 657 peaks common only to them, Pdx1 and NeuroD1 had 2225 common only to them, and strikingly, there were 12508 peaks commonly bound by only NeuroD1 and MafA. There were 7470 peaks with all three transcription factors bound. Again, this suggests NeuroD1 and MafA have very similar binding patterns, as evidenced by their similarities in their identified motifs (Figure 22).

To look at genes that were dually or triply regulated by the transcription factors, I performed RNAi of combinations of transcription factors, i.e. Pdx1 and NeuroD1, Pdx1 and MafA, NeuroD1 and MafA, and Pdx1 NeuroD1 and MafA in biological quadruplicates. After 48 hours of RNAi, the expression of Pdx1 and NeuroD1 decreased to 30 and 43% respectively, the expression of *Pdx1* and *MafA* decreased to 40% and 20% respectively, and *NeuroD1* and *MafA* levels decreased to 40% and 15% respectively (Figure 26). The knockdown efficiency was reduced in the triple transfection efficiency despite optimization (data not shown), with *Pdx1*, *NeuroD1* and *MafA* mRNA levels remaining at 40-55% after 48 hours of RNAi. In order to assess global gene expression changes in the RNAi samples, Illumina bead chip arrays were used. In response to the depletion of endogenous *Pdx1* and *NeuroD1*, 1167 genes were significantly regulated ±1.5 fold, $p<0.05$. In response to the depletion of *Pdx1* and *MafA*, 2339 genes were significantly regulated ±1.5 fold, $p<0.05$. When *NeuroD1* and *MafA* mRNA levels were reduced, 1424 genes were significantly regulated ±1.5 fold, $p<0.05$. A complete list of genes in each of the microarray datasets is available in Supplementary Table 5 and 6.

To look at how Pdx1, NeuroD1 and MafA genes were co-regulated, genes that were changing upon each combination of transcription factor knockdown were overlapped (Figure 27) and compared. Strikingly, 1427 genes were regulated by the Pdx1-MafA only combination, in comparison to the 242 genes and 499 genes regulated by the Pdx1-NeuroD1 and NeuroD1-MafA combination respectively.

(A)



(B)

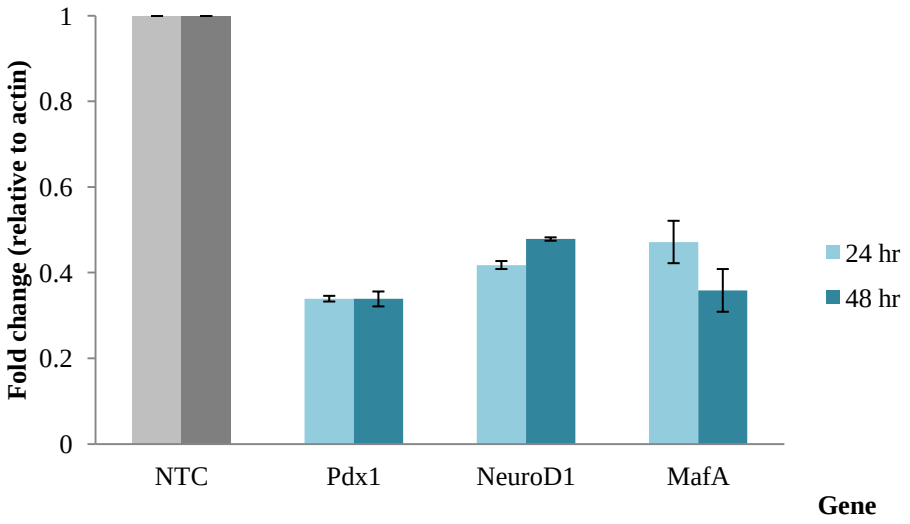


Figure 26. siRNA depletion of endogenous *Pdx1*, *NeuroD1* and *MafA* mRNA in combination with one another. (A) *Pdx1*, *NeuroD1* and *MafA* levels were depleted by RNA interference using siRNA against eqimolar concentrations of *Pdx1* and *NeuroD1* siRNA, *Pdx1* and *MafA* siRNA, and *NeuroD1* and *MafA* siRNA in NIT-1 cells. RNA was harvested 48hrs post-transfection and transcript levels assayed by RT-PCR. Mean values ± S.E. are plotted as percentages relative to NTC (100%). The siRNA experiments were performed in quadruplicate and normalized to endogenous *β-actin*. (B) *Pdx1*, *NeuroD1* and *MafA* were simultaneously depleted in equimolar concentrations by RNA interference. RNA was harvested 24 and 48hrs post-transfection and transcript levels assayed by RT-PCR. Mean values ± S.E. are plotted as percentages relative to NTC (100%). The siRNA experiments were performed in quadruplicate and normalized to endogenous *β-actin*.

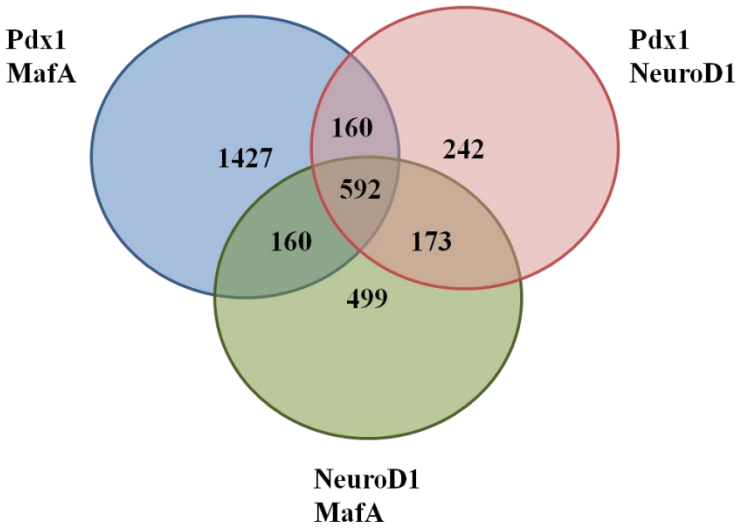


Figure 27. Overlap of Pdx1-MafA, Pdx1-NeuroD1 and NeuroD1-MafA regulated genes. Pdx1, NeuroD1 and MafA RNAi genes regulated ±1.5 fold show significant overlap. (Gireesh Bogu performed the analysis.)

For ease of explanation in this manuscript, I use the term "regulated" for genes that changed expression (up- or down-regulated) in response to a transcription factor knockdown, while genes that show the binding of a transcription factor are termed as "bound". To explore how combinatorial transcription factor binding correlated with changes in gene expression upon combinatorial transcription factor knockdown, the number of regulated genes (i.e. genes that change in response to single or double or triple knockdown of the transcription factors) were overlapped with genes that show transcription factor binding (i.e. the gene may be bound by one, two or all three transcription factors).

To achieve that, transcription factor ChIP-Seq peaks were resized to 65bp from peak center and peaks that were within ±2kb distance from a TSS were selected for this analysis. Genes with a Pdx1 binding site (00P) show the most change in expression in the single transcription factor knockdown experiments, be it Pdx1-depleted cells or NeuroD1-depleted cells or MafA-depleted cells (Figure 28A). There are also more genes going up upon each transcription factor knockdown, regardless of the transcription factor that is depleted. This suggests that in general, the three transcription factors are acting as repressors, rather than activators. A similar trend is seen with genes that have a NeuroD1 binding site (00N), but the number of genes changing expression in this group is less than those with a Pdx1 binding site. When there are NeuroD1 and MafA binding sites present (0MN), there are more genes changing in the NeuroD1-knockdown experiment. The presence of MafA alone (00M) appears to affect the regulation of genes to a smaller extent.

The same gene sets showing combinations of transcription factor binding sites as in Figure 28A were overlapped with genes found to be regulated by the double transcription factor knockdowns (Figure 28B). Similar to what was seen in Figure 28A, there are the most number of genes changing in all the double transcription factor knockdown experiments, be it Pdx1-NeuroD1-depletion or Pdx1-MafA-depletion or NeuroD1-MafA-depletion, as long as there is a Pdx1-binding site present (00P). Again, there are more genes are up-regulated instead of

down-regulated. There are also more number of genes changing in the double transcription factor knockdown that have a Pdx1 binding site (100 - 250) as compared to 50 – 170 in the single transcription factor knockdown. Genes with both NeuroD1 and MafA binding sites (0MN) and those with just NeuroD1 binding sites (00N) also showed an increase in the number of genes that they were regulating (50-100), when compared to 0MN and 00N in the single transcription factor knockdown (10-50 genes). The similarties in behavior of 00N and 0MN could be partly due to the similarities in the binding sites of NeuroD1 and MafA, as previously identified (Figure 22).

When genes bound by transcription factor binding sites were overlapped with genes regulated by the triple transcription factor knockdown (Figure 28C), genes with a Pdx1 binding site (00P) showed the most number of genes changing, as observed previously. The number of genes that changed were fewer than that observed in the double knockdown (130 up- and down-regulated, compared to 100 – 250 up- and down-regulated), perhaps due to the lesser degree of efficiency in the knockdown levels of Pdx1, NeuroD1 and MafA in the triple knockdown experiment (Figure 28B).

Taken together, Figure 28 shows that genes containing Pdx1 binding sites in their promoters display maximal differential expression in the single, double or triple knockdown experiments. There are also more genes going up upon each transcription factor knockdown, regardless of the transcription factor that is depleted, suggesting that these transcription factors act more as repressors than activators.

To tease apart the relationship between genes that are bound and regulated by the transcription factor combinations, the fraction of bound genes [termed B, i.e. genes that show transcription factor binding in the gene promoter (± 2kb from TSS)] that are regulated [i.e. genes that are bound by a transcription factor and show differential expression upon transcription factor knockdown (termed BR)] was expressed as a percentage (i.e. BR*100/B), and plotted as a

heatmap (Figure 29). This was to gain insight into the genes that are most responsive to transcription factor binding. The x-axis indicates the number of genes regulated, as identified in the microarray data from siRNA-depleted transcription factors. The y-axis indicates transcription factor binding sites (TFBS) in various combinations. The colors indicate percentage of genes that are bound and regulated (BR), as a fraction of the total number of bound genes (BR/B), with red being the most (10%), yellow being in the middle (6-8%) and blue being the least (<4%). There were three distinct clusters of genes: the Pdx1-MafA, Pdx1-NeuroD1, and NeuroD1-MafA cluster (red), the NeuroD1 and Pdx1-NeuroD1-MafA regulated gene cluster (yellow), and the Pdx1 and MafA alone regulated gene cluster (blue). Genes that are triply and doubly bound have the highest percentage of genes that are regulated by the double transcription factor knockdown (red cluster), while a lesser percentage of genes are regulated by NeuroD1 or all three transcription factors (yellow cluster). The Pdx1 and MafA single transcription factor knockdowns have the lowest percentage of BR genes (blue cluster). This indicates that the triply bound genes have the highest effect on regulated genes, compared to the doubly and singly bound genes.

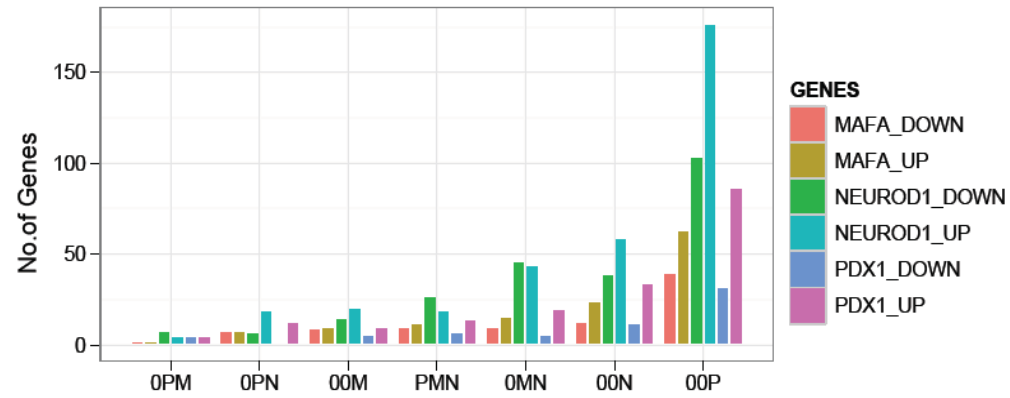
The highest percentage of genes regulated is those in the double knockdown, irrespective of whether the gene is bound by single, double or all three transcription factors. The fact that the doubly and triply bound genes show a high BR/B value is not merely an artifact of the lesser number of doubly and triply bound genes (i.e. lesser number of genes in the denominator (B) would inflate the value of BR/B). If that were the case, only the triply bound genes would show the highest percentage of BR/B, because the number of genes triply bound is the fewest of all the bound combinations of genes. However, the data shows that a high percentage of BR is also observed in double knockdown genes that are doubly and singly bound.

Of the single transcription factor regulated genes, NeuroD1-regulated genes had a higher percentage of bound genes (6-8%) compared to Pdx1 and MafA (<4%). This could be due to the fact that 12.5% of NeuroD1-regulated genes were transcription factors (Table 6), as

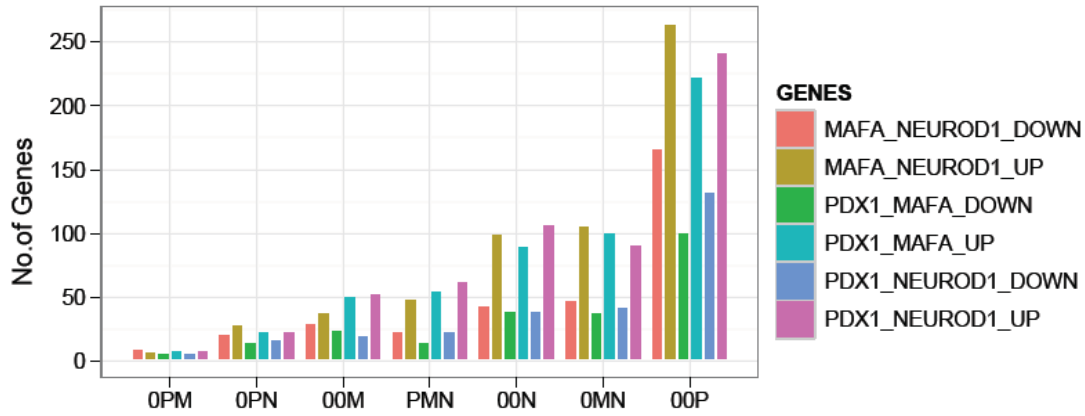
compared to 8.5% of Pdx1 and MafA target genes. This higher proportion of transcription factors in NeuroD1-regulated genes suggests that NeuroD1 mediates downstream gene expression changes via other transcription factors. The effect of the genes regulated by the triple knockdown was less than those regulated by the double knockdown (Figure 29). This is perhaps due to the lesser degree of efficiency in the knockdown levels of *Pdx1*, *NeuroD1* and *MafA* in the triple knockdown experiment.

Taken together, this figure shows that triply bound genes have the most effect on gene regulation, followed by genes that are doubly bound by transcription factors. Genes bound by a single transcription factor have a fewer percentage of regulated genes.

(A)



(B)



(C)

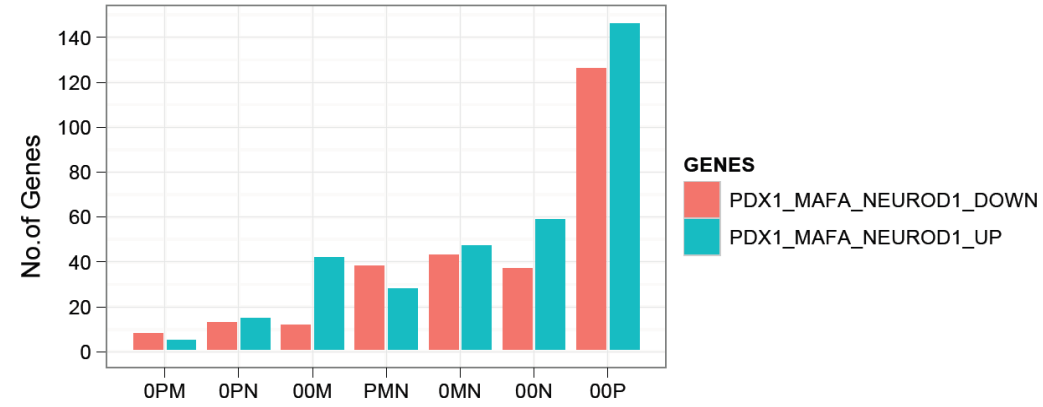


Figure 28. Genes containing Pdx1 binding sites have the most effect on regulated genes. Transcription factor peaks at the nearest gene promoter (±2kb from TSS) were resized to 65bp peaks (±65 bp from peak center) and overlapped with genes identified to be regulated in single (A), double (B) and triple (C) transcription factor knockdowns. 00P: Pdx1 binding sites, 00M: MafA binding sites, 00N: NeuroD1 binding sites; 0PM: Pdx1 and MafA binding sites, 0PN: Pdx1 and NeuroD1 binding sites, 0MN: MafA and NeuroD1 binding sites; PMN: Pdx1, MafA and NeuroD1 binding sites. y axis = number of genes regulated, as identified in the microarray of siRNA-depleted transcription factor, x axis = transcription factor binding sites. Genes containing Pdx1 binding sites in their promoters display maximal differential expression in the single, double or triple knockdown experiments. There are also more genes going up upon each transcription factor knockdown, regardless of the transcription factor that is depleted, suggesting that these transcription factors act more as repressors than activators. (Gireesh Bogu performed the analysis.)

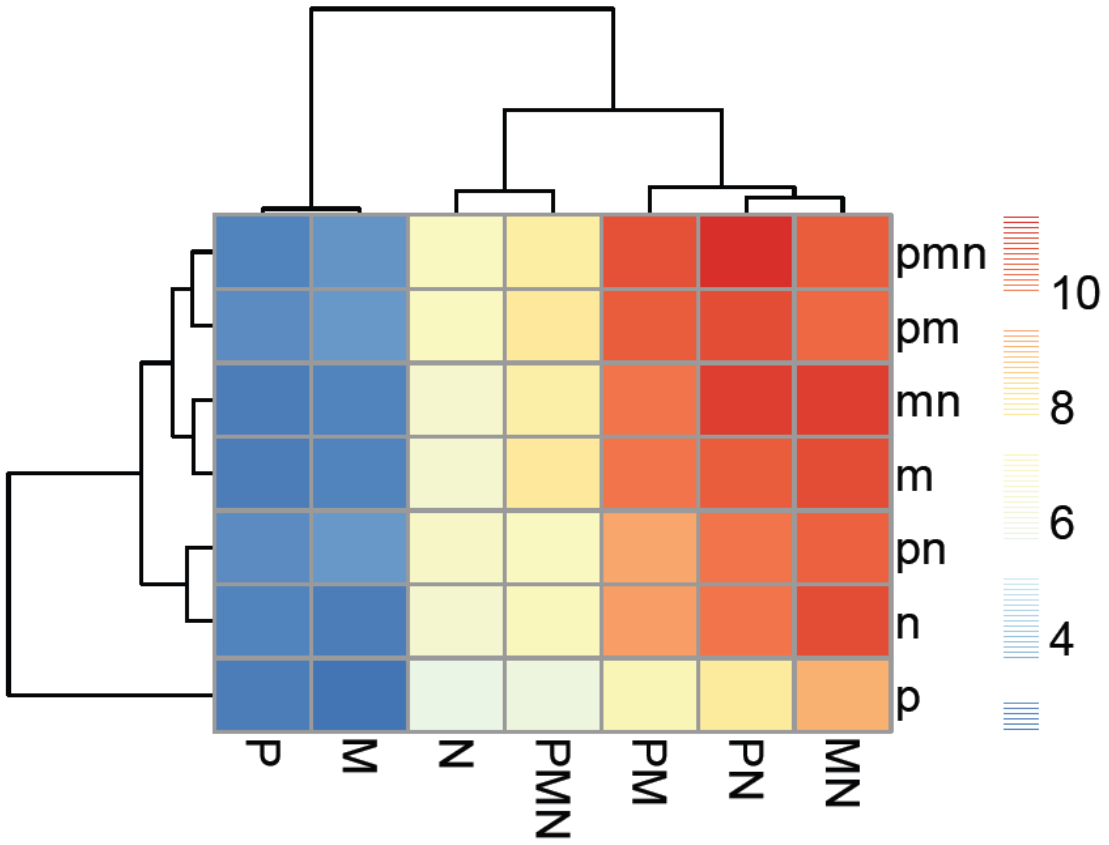


Figure 29. Triply bound genes have the highest effect on regulated genes, compared to the doubly and singly bound genes. The fraction of bound genes [termed B, i.e. genes that show transcription factor binding in the gene promoter (± 2kb from TSS)] that are regulated [i.e. genes that are bound by a transcription factor and show differential expression upon transcription factor knockdown (termed BR)] was expressed as a percentage (i.e. BR*100/B), and plotted as a heatmap. The x-axis indicates the number of genes regulated, as identified in the microarray of siRNA-depleted transcription factor. The y-axis indicates transcription factor binding sites in various combinations. The colors indicate percentage of genes that are bound and regulated (BR), as a fraction of the total number of bound genes (BR/B), with red being the most (10%), yellow being in the middle (6-8%) and blue being the least (<4%). Triply bound genes have the most effect on gene regulation, followed by genes that are doubly bound by transcription factors. Genes bound by a single transcription factor have a fewer percentage of regulated genes. (Dr Akshay Bhinge suggested this analysis and Gireesh Bogu performed it.)

3.3. DISCUSSION

3.3.1. THE PDX1, NEUROD1 AND MAFA GENOMIC BINDING LANDSCAPE

Pdx1 and NeuroD1 co-occupy 440 genomic loci[86], suggesting that a significant fraction of Pdx-1 and NeuroD1 binding sites exist within regulatory modules, where they may control gene expression in a synergistic manner. However, it was unknown whether MafA interacts at similar loci to co-regulate genes important for β-cell function, as no genome-wide binding studies have been performed for MafA. Besides, the Pdx1 and NeuroD1 experiments were carried out on custom-made promoter arrays which contained promoter elements, enhancers and highly conserved regions, genomic elements that represent miRNAs, and putative regulatory elements identified through serial analysis of chromatin occupancy (SACO). Therefore, there is bias in the data due to the pre-selection of putative binding sites, repeat masking of the arrays, and less sensitivity due to a lack in base-pair resolution of binding sites. With new generation ChIP-Seq technology, these shortcomings are overcome as sensitivity and statistical certainty can be adjusted to detect rare DNA-protein interaction sites. The accuracy is also not limited by the spacing of pre-determined probes, and binding affinities of a protein to different regions of DNA can be compared[213]. As such, to address my hypothesis that Pdx1, NeuroD1 and MafA were bound at similar and dissimilar sites across the genome, I performed ChIP-Seq with chromatin isolated from the β-cell line NIT-1, with Pdx1, NeuroD1 and MafA antibodies.

23132 peak regions were identified to be Pdx1-bound, 31031 peak regions NeuroD1-bound and 24108 peak regions MafA-bound (Table 2). Pdx1 shows very high relative enrichment in promoters, with 17.3% of ChIP regions enriched at the promoter compared 6.4% in NeuroD1 and 5.7% in MafA (Figure 19). MafA regulates the Pdx1 promoter in β-cells[122], and it was not previously known if other co-regulatory process occur. I have shown that Pdx1, NeuroD1 and MafA bound to one another's promoters, along with the *Insulin* promoter (Figure 15). The

Pdx1, NeuroD1 and MafA peaks are enriched around all three peak centers, suggesting they associated strongly with one another (Figure 21).

The transcription factor motifs found to be enriched by HOMER (Figure 22) demonstrated that the ChIPs were robust and of good quality. HOMER is a *de novo* motif discovery algorithm that scores motifs by looking for motifs with differential enrichment between two sets of sequences, the first being the target sequences of interest (i.e. the ChIP-Seq dataset) and the second being a set of background sequences (i.e. whole genome background) [189]. *De novo* motif discovery is a more unbiased method of motif discovery than known motif scanning, which involves finding all the known motifs in a sequence. The distinction between our *de novo* motif discovery and traditional *de novo* motif discovery is that in the former, all the ChIP-Seq peaks are input into HOMER, and in the latter, the top several hundred ChIP-Seq peaks with highest p values are used. Therefore, our method of *de novo* motif discovery is a more stringent and robust method of identifying motifs enriched in the dataset. Enrichment values reported by HOMER should be extremely significant, $<< 1^{e-50}$, and the p values of these motifs I identified were far more significant.

The known Pdx1 motif was identified in my Pdx1 ChIP-Seq dataset, with a p value of $1e^{-260}$. The motifs for NeuroD1 and MafA were almost identical, with highly significant p values of $1e^{-3395}$ and $1e^{-3104}$ respectively. It is noteworthy that they had such similarities, probably due to the large overlaps in the number of binding sites (Figure 23). Interestingly, the motif for the transcription factor Atoh1 came up on top (p value $1e^{-3395}$). Like NeuroD1, Atoh1 is a bHLH transcription factor that activates E-box dependent transcription. Apart from its role in endodermal development and pancreatic function, Neurod1 is also crucial for mediating the differentiation of cerebellar granule cells downstream of Atoh1-mediated proliferation of the precursors[214] and has recently been shown to be a direct target of Atoh1 in the developing brain[215]. Another motif that was highly enriched in the NeuroD1 and MafA ChIP-Seq datasets is E2A, with p values of $1e^{-554}$ and $1e^{-547}$ respectively. The E2A gene encodes transcription

factors of the bHLH family which are implicated in cell-specific transcriptional control in several cell lineages. The second domain (located between amino acids 369 and 485) functions preferentially in pancreatic β-cell lines, showing a pattern of heptad repeats of leucine residues characteristic of bLZ transcription factors[216]. Therefore, the selective properties of this activation domain may contribute to cell-specific transcription directed by the E2A gene. NeuroD1 itself is unable to bind DNA and functions in a complex with the more generally distributed E2A (i.e. E47, E12, E2-5)-encoded proteins[217], and the presence of a tissue-enriched and a ubiquitously distributed bHLH factor in the activator complex is characteristic of other tissue-specific members of this family, the best characterized of which are involved in myogenic[218] and neuronal differentiation[219]. These activators bind to the consensus sequence CANNTG (E-box), with hetero-dimerization increasing DNA binding and activation capacity. NeuroD1-directed transcription of *Insulin* has been shown to occur via the E-box within the *Insulin* promoter[97,107]. Taken together, these results confirm that majority of the binding sequences are relevant for Pdx1, NeuroD1 and MafA.

3.3.2. Pdx1, NeuroD1 and MafA Regulate Biologically Similar Networks

There were many more Pdx1-only peaks (12780) when compared to NeuroD1 and MafA (8828 and 3743) (Figure 23). Strikingly, 12508 peaks were bound by both MafA and NeuroD1, compared to 657 peaks common to both Pdx1 and MafA and 2225 to Pdx1 and NeuroD1. This suggests that Pdx1 is acting more independently, with MafA and NeuroD1 having very similar and possibly redundant roles. A more independent role of Pdx1 would also concur with the fact that 17% of Pdx1 peaks were found in promoter regions, compared to NeuroD1 and MafA. When Pdx1, NeuroD1 and MafA were depleted, most genes were changing upon NeuroD1 depletion (1550), in comparison to Pdx1 (570) and MafA (522). To look at how Pdx1, NeuroD1 and MafA genes were co-regulated, genes that were changing upon each transcription factor knockdown were overlapped and compared (Figure 25). In line with the fact that depletion of NeuroD1 had more effect on target genes, 834 genes were

regulated by NeuroD1 only, in comparison to the 175 and 151 regulated by Pdx1 and MafA respectively. This suggests that NeuroD1 is acting in an indirect manner. In all three cases, there was a greater number of genes being up-regulated upon knock-down, indicating that these transcription factors had a generally more repressive role on their target genes, rather than activating.

To shed biological insight into the network of genes bound and regulated by Pdx1, NeuroD1 and MafA, the ChIP-Seq peaks and siRNA-microarray genes were uploaded into IPA and PANTHER. Genes targeted by Pdx1, NeuroD1 and MafA were over-represented in highly significant pathways including those involved in metabolic processes (e.g. protein metabolism, nucleoside metabolism, lipid metabolism), developmental processes (e.g. response to pheromones, organ and nervous system development), and cellular processes (e.g. signal transduction, vesicle-mediated transport, cell cycle, cell death, cellular growth and proliferation) (Tables 3-5). These annonations suggest that Pdx1, NeuroD1 and MafA all interact in a similar fashion to common regulatory roles in maintaining the β-cell phenotype, by means of regulating metabolic activities, organ development and cell cycle activities.

3.3.3. Combinatorial Binding of Pdx1, NeuroD1 and MafA

As the binding sites of Pdx1, NeuroD1 and MafA and the genes they regulated categorized into similar regulatory networks, and these three transcription factors synergistically bind to and regulate the *Insulin* promoter[124], I hypothesized that these three transcriptional factors would combinatorially regulate other genes in the genome. To explore how combinatorial transcription factor binding correlated with changes in gene expression upon combinatorial transcription factor knockdown, the number of regulated genes that had any combination of a transcription factor was overlapped with the genes identified to be changing upon each combination of transcription factor knockdown. Genes containing Pdx1 binding sites in their promoters display maximal differential expression in the single, double or triple knockdown

experiments (Figure 28). There are also more genes going up upon each transcription factor knockdown, regardless of the transcription factor that is depleted, suggesting that these transcription factors act more as repressors than activators. The highest percentage of genes regulated is those in the double knockdown, irrespective of whether the gene is bound by single, double or all three transcription factors (Figure 29). Of the single transcription factor regulated genes, NeuroD1-regulated genes had a higher percentage of bound genes (6-8%) compared to Pdx1 and MafA (<4%). This could be due to the fact that 12.5% of NeuroD1-regulated genes were transcription factors (Table 6), as compared to 8.5% of Pdx1 and MafA target genes. This higher proportion of transcription factors in NeuroD1 regulated genes suggests that NeuroD1 mediates downstream gene expression changes via other transcription factors. Taken together, these figures show that while Pdx1 on its own had the most effect on gene regulation, triply and doubly bound genes have more effect on gene regulation.

3.4. SUMMARY

There is increasing evidence that transcription factors act synergistically to achieve normal pancreatic development and function. Rapid progress has been made to elucidate regulatory mechanisms controlled by key β-cell genes. In this chapter, I have looked at the transcription network regulated by Pdx1, NeuroD1 and MafA. I have shown that they are involved in regulating metabolic processes, developmental processes and cellular processes. Of the three transcription factors, Pdx1 appears to be the dominant one, while NeuroD1 has more of an indirect role in regulating β-cell genes. When the three transcription factors are bound at the same gene, there is most effect on gene regulation. Improving our knowledge of these important transcription factors, establishing their hierarchy, identifying and characterizing novel transcription factors, and ultimately regulating their expression to ensure glucose homeostasis may be the key to prevent or correct β-cell dysfunction.

CHAPTER 4:

PDX1, NEUROD1 & MAFA TARGET GENES HAVE FUNCTIONAL EFFECTS ON THE BETA-CELL PHENOTYPE

4.1. INTRODUCTION

Transcription factors play an important role in pancreatic β-cell development and function. This is manifested clinically by the MODY syndromes, in which heterozygous mutations in HNF4α, HNF1α, PDX1, HNF1β and NEUROD1 cause progressive impairments in insulin secretion and the eventual development of diabetes mellitus[90]. To understand how β-cell transcription factors might regulate β-cell function, it is often useful to think of these factors as forming a network that ultimately activates a set of unique β-cell genes (e.g. *Ins*, *Glut2*, *Iapp* or *Gck*). The main function of the mature β-cell is to respond appropriately with the secretion of insulin, to the fluctuating levels of blood glucose. This is mediated and controlled by many factors, of which Pdx1, NeuroD1 and MafA are a few.

Pdx1 mediates the adaptive responsiveness of β-cells to changes in extra-cellular glucose concentrations by regulating expression levels of glucose sensors, such as the aforementioned *Glut2*[91] and *Gk*[92], and through the activation of glucose responsive enhancers within the Insulin promoter, in conjunction with other transcriptional functions. It also controls the expression of *Syt1*, *Nkx6.1*[94], itself[86], and also functions downstream of the insulin signaling pathway in the regulation of β-cell mass[85]. However, the number of target genes is certainly much greater, as *Pdx1*$^{-/-}$ mice display pleiotropic phenotypes[95]. NeuroD1 is implicated in the regulation of insulin secretion via modulation of the expression of ion channel transporters and channels[86] e.g. *Nna*[113] and *Sur1*, which forms potassium channels with Kir6.2[111]. The mis-expression of NeuroD1 in a human fetal epithelial cell line induced the expression of several genes required for vesicular trafficking and exocytosis, including *Sec24D*, *Snap25*, *Stx1* and *Munc18*[114]. This finding suggests that NeuroD1 may regulate insulin exocytosis in pancreatic β-cells by directly inducing the expression of genes involved in exocytosis. In contrast to Pdx1 and NeuroD1, β-cell development does not appear to be affected in MafA deficient mice, although adult mutant animals are glucose intolerant due to diminished *Insulin* transcription and impaired GSIS, decreased β-cell number and altered islet architecture[43].

MafA appears to also play a role in mediating the expression of *grauphilin*, *adiponectin*, *Glut2*, *Nkx6.1*, *Pdx1*[122], *pyruvate carboxylase*, and *prohormone convertase 1/3*[32].

As Pdx1, NeuroD1 and MafA have been shown to co-regulate many genes of similar biological functions (Chapter 3), **I hypothesized that the genes bound to and regulated by all three transcription factors would play an important role in the main function of the mature β-cell, which is to secrete insulin in response to glucose.** To address this question, I decided to study in greater detail 16 genes that were: (1) Bound by all three transcription factors and (2) had a change in gene expression upon depletion of each transcription factor. I hypothesized that, if these Pdx1 NeuroD1 and MafA target genes were important in maintaining the insulin secretory response, depletion of their expression in β-cells would result in impaired ability to respond to glucose, as determined by a GSIS assay. I also looked at changes in expression of critical genes involved in the GSIS response to better understand the molecular mechanisms behind the effects of these genes.

4.2. RESULTS

4.2.1. PDX1, NEUROD1 AND MAFA REGULATE THEIR TARGET GENES

In order to test my hypothesis that downstream target genes of Pdx1, NeuroD1 and MafA are of functional importance to the mature β-cell, I sought to test the effects of their depletion on NIT-1 cells. I selected 16 genes (Table 8) from an overlap of genes commonly bound by Pdx1, NeuroD1 and MafA, that also had expression changes upon each transcription factor knockdown in the microarray. Most of these genes had no previous known function or role in the pancreatic β-cell, and have not been well characterized in other cell types. They are: *Extracellular leucine-rich repeat and fibronectin type III domain containing (Elfn)-2, Exocyst complex component (Exoc)-4, Integrin β 1 binding protein (melusin) (Itgb1bp)-2, Glucosaminyl (N-acetyl) transferase 2 I-branching enzyme (I blood group) (Gcnt)-2, Salt-inducible kinase (Sik)-2, Protein phosphatase 1 regulatory (inhibitor) subunit (Ppp1r)-3a, Septin7, Aldolase B fructose-bisphosphate (Aldob), Histone cluster 1 (Hist1)-H1c, ATPase Na+/K+ transporting β1 polypeptide (Atp1b1), Solute carrier family 1 (glial high affinity glutamate transporter) member 2 (Slc1a2), Complexin (Cplx)-2, Coagulation factor XIII A1 polypeptide (F13a1), Src homology 2 domain containing transforming protein (Shc)-3, TSPY-like (Tspyl)-4*, and *BEN domain containing (Bend)-7*.

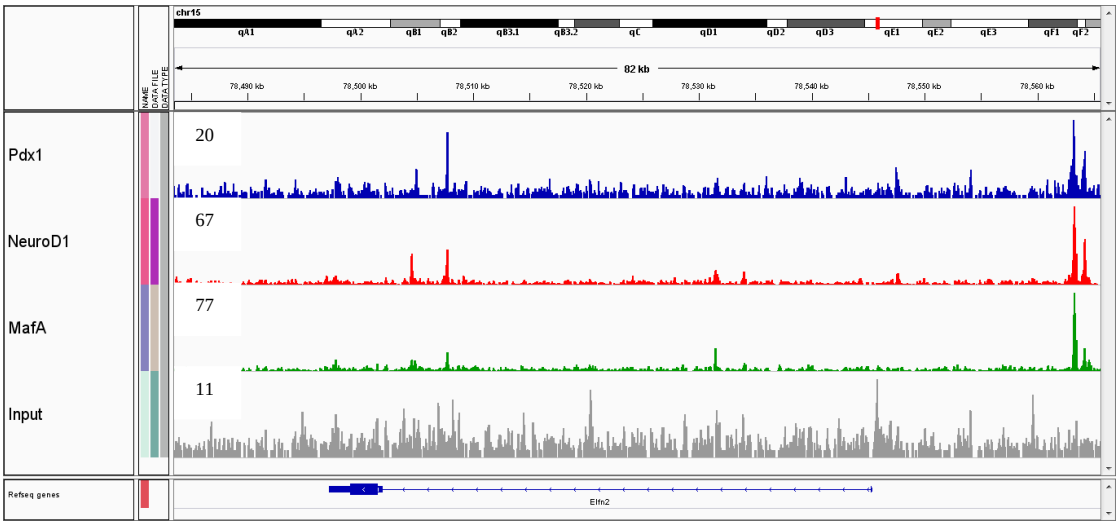
I verified they were bound by the transcription factors and had changes in gene expression in response to each transcription factor knockdown. Figure 30 and the heatmap in Figure 31, shows they were all bound by Pdx1, NeuroD1 and MafA relative to the control input library, and all had either increases or decreases in gene expression both on the microarray and by qPCR validation. E.g. in Figure 30A, Pdx1, NeuroD1 and MafA have peaks approximately 1kb upstream of the *Elfn2* TSS, with peak heights of 20, 67 and 77 respective to the input control library which has non-specific peaks of 11 in that same region. These results together confirm that the target regions selected for functional characterization are indeed occupied and regulated individually by Pdx1, NeuroD1 and MafA in the mouse NIT-1 β-cell line.

Gene	Bound by Pdx1 NeuroD1 and MafA, with ↑ /↓ in each TF siRNA (log2)	Function
Elfn2	↑ 1.55	Expressed in pancreas. Nothing known about it.
Exoc4	↑ 1.73	A multiple protein complex essential for targeting exocytic vesicles to specific docking sites on the plasma membrane[220]. Polymorphisms identified in T2DM patients[221].
Itgb1bp2	↑ 2.13	Lowly expressed in pancreas. May play a role during maturation and/or organization of muscle cells[222].
Gcnt2	↑ 2.03	Closely associated with the development and maturation of erythroid cells[223]. Identified as a down-regulated target of miRs expressed highly in pancreatic islet-like cell clusters differentiated from human embryonic stem cells[224].
Sik2	↑1.53	Phosphorylates Ser-794 of IRS1 in insulin-stimulated adipocytes, potentially modulating the efficiency of insulin signal transduction[225].
Ppp1r3a	↑1.53	Plays an important role in glycogen synthesis but is not essential for insulin activation of glycogen synthase[226]. A common variant is associated with insulin resistance and T2DM[227].
Septin7	↓ 1.53	E2F1/2 responsive genes in the pancreas, including Septin 2 and 7, are typically regulated at G2/M in the cell cycle[228]. Filament-forming cytoskeletal GTPase and required for normal organization of the actin cytoskeleton[229]. Involved in cytokinesis[230].
Aldob	↑1.31	Glycolytic enzyme that catalyzes the reversible conversion of fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate[231]. Highly expressed in liver. Regulated by HNF4α in the pancreas[232].
Hist1H1c	↓ 1.15	Histones H1 are necessary for the condensation of nucleosome chains into higher order structures[233].
Atp1b1	↑1.53	This is the non-catalytic component of Na+/K+ -ATPase, which catalyzes the hydrolysis of ATP coupled with the exchange of Na(+) and K(+) ions across the plasma membrane. The β subunit regulates, through assembly of α/β heterodimers, the number of sodium pumps transported to the plasma membrane[234].

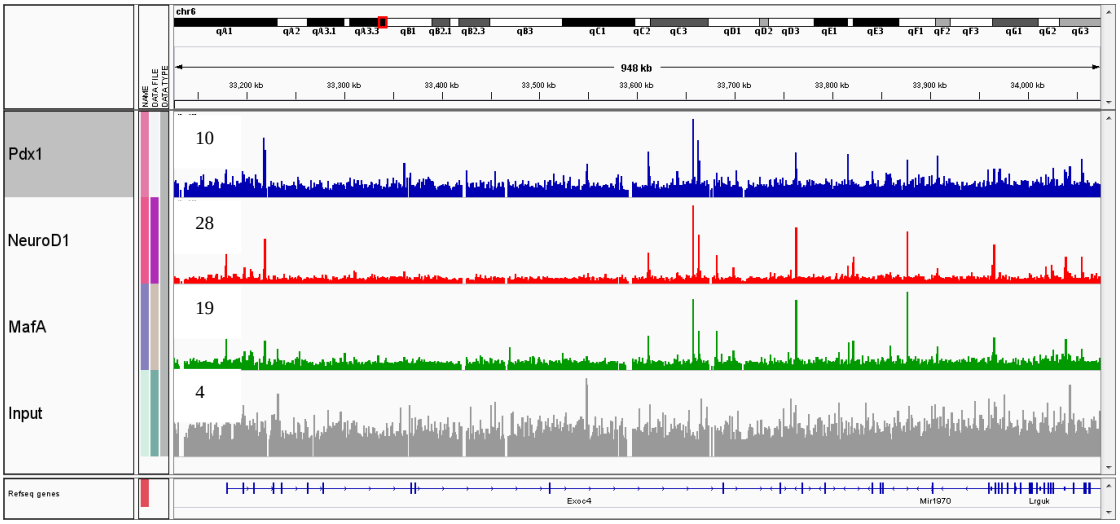
Gene	Bound by Pdx1 NeuroD1 and MafA, with ↑ /↓ in each TF siRNA (log2)	Function
Slc1a2	↓ 1.44	Transports L-glutamate, L- and D-aspartate, essential for terminating the postsynaptic action of glutamate by rapidly removing released glutamate from the synaptic cleft[235]. Expressed in pancreatic β-cells and prevents glutamate-induced β-cell death[236].
Cplx2	↑1.73	Positively regulates a late step in synaptic vesicle exocytosis[237].
F13a1	↓ 1.68	Encodes the coagulation factor XIII A subunit, which is the last zymogen to become activated in the blood coagulation cascade[238].
Shc3	↓ 1.84	Expressed in brain and pancreas. Signaling adapter that couples activated growth factor receptors to signaling pathway in neurons[239].
Tspyl4	↓ 1.9	Nothing known about it.
Bend7	↓ 1.37	Nothing known about it.

Table 8. Genes bound and regulated by Pdx1, NeuroD1 and MafA**.**

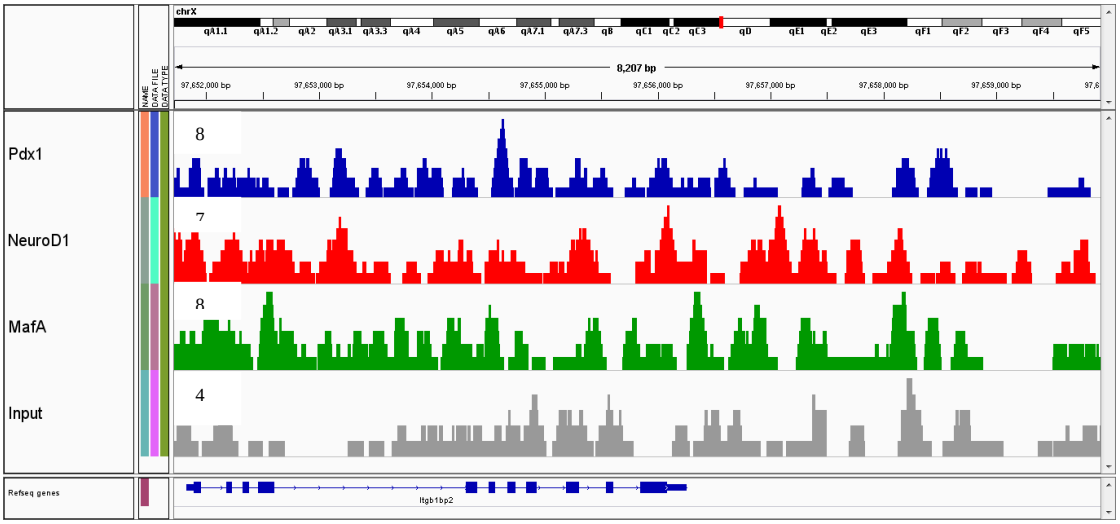
(A) Elfn2



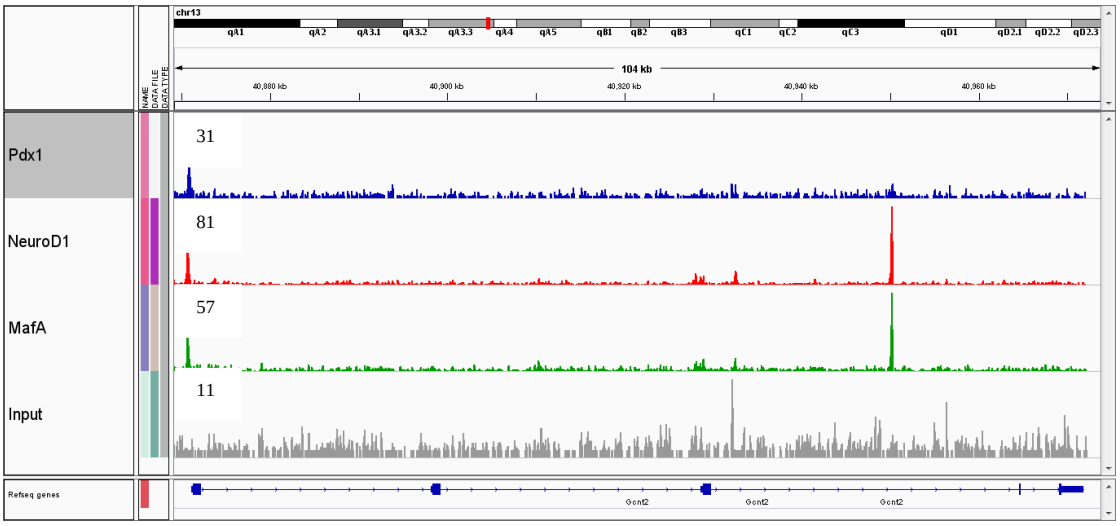
(B) Exoc4



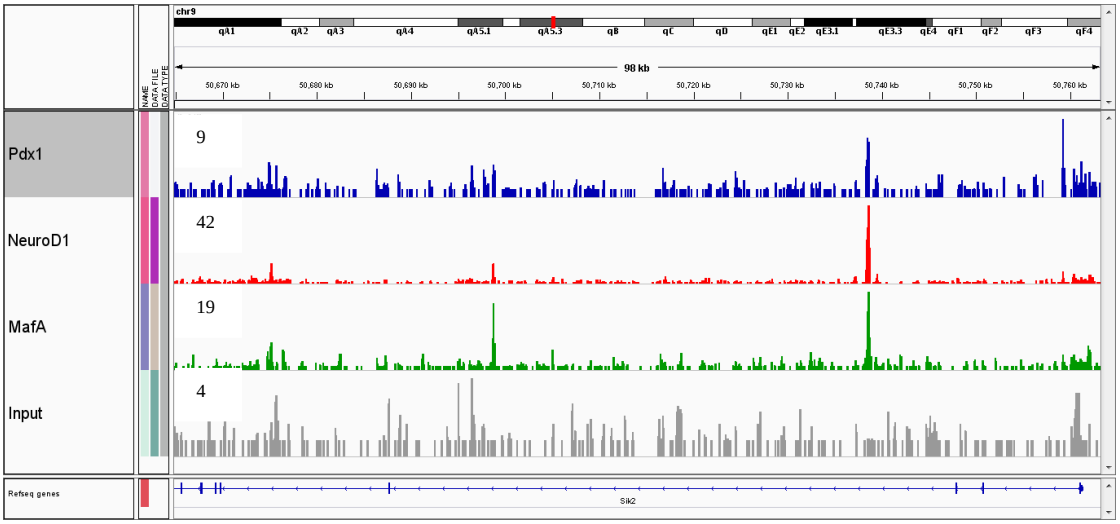
(C) Itgb1bp2



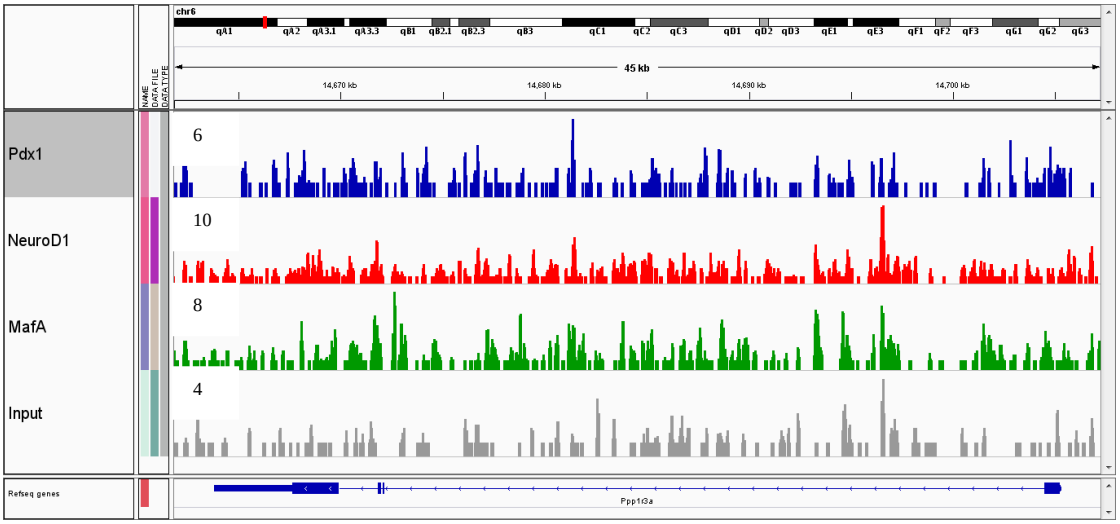
(D) Gcnt2



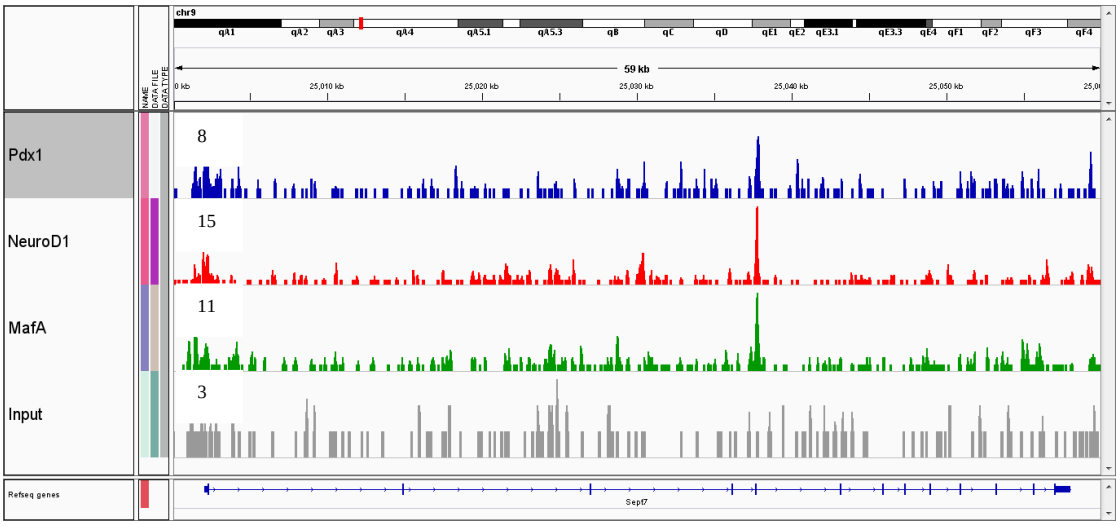
(E) Sik2



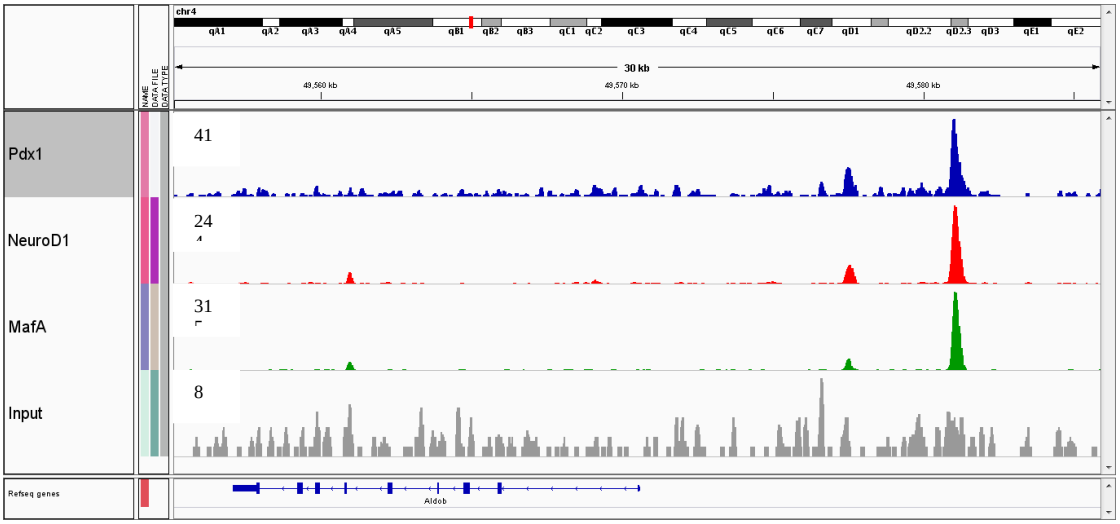
(F) Ppp1r3a



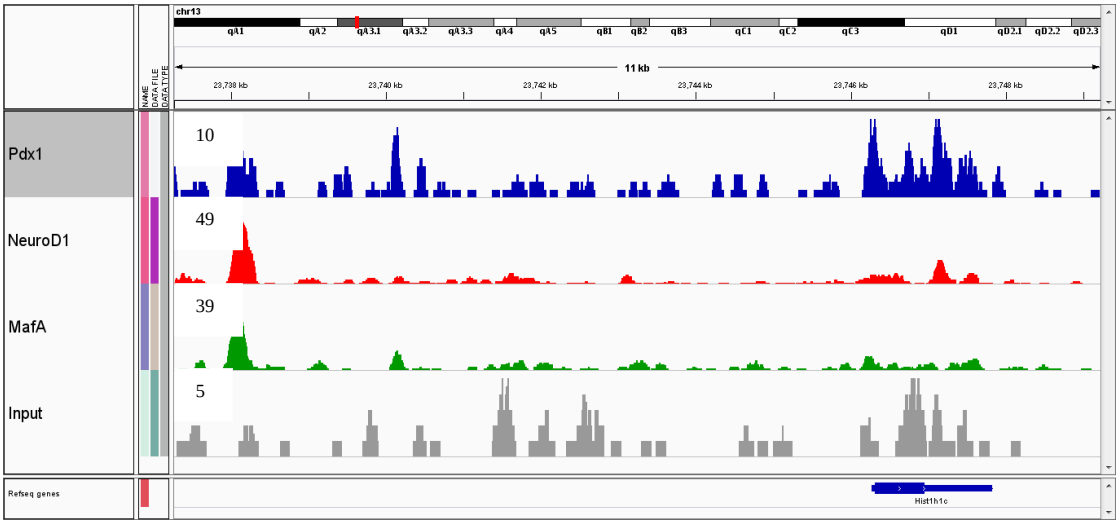
(G) Septin7



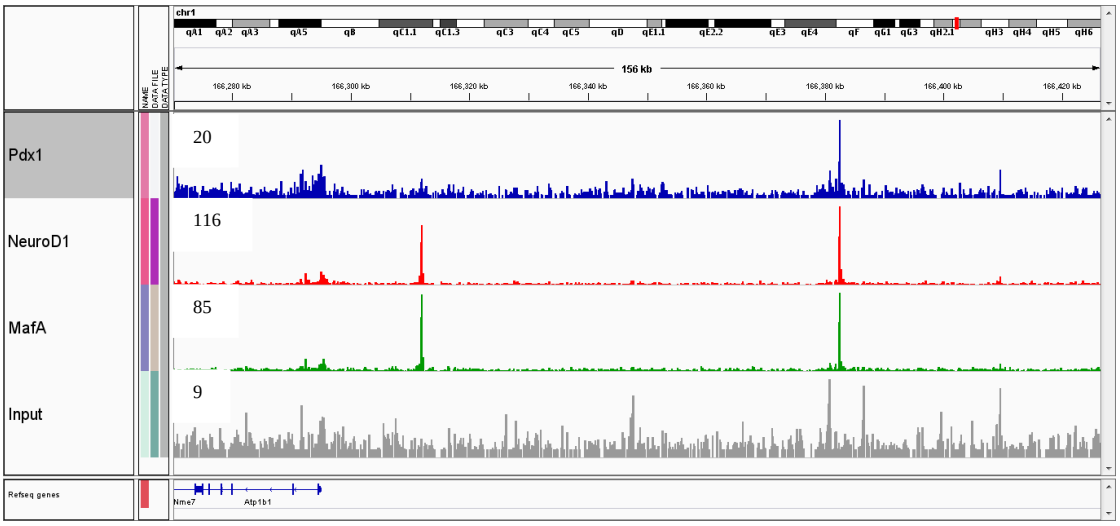
(H) Aldob



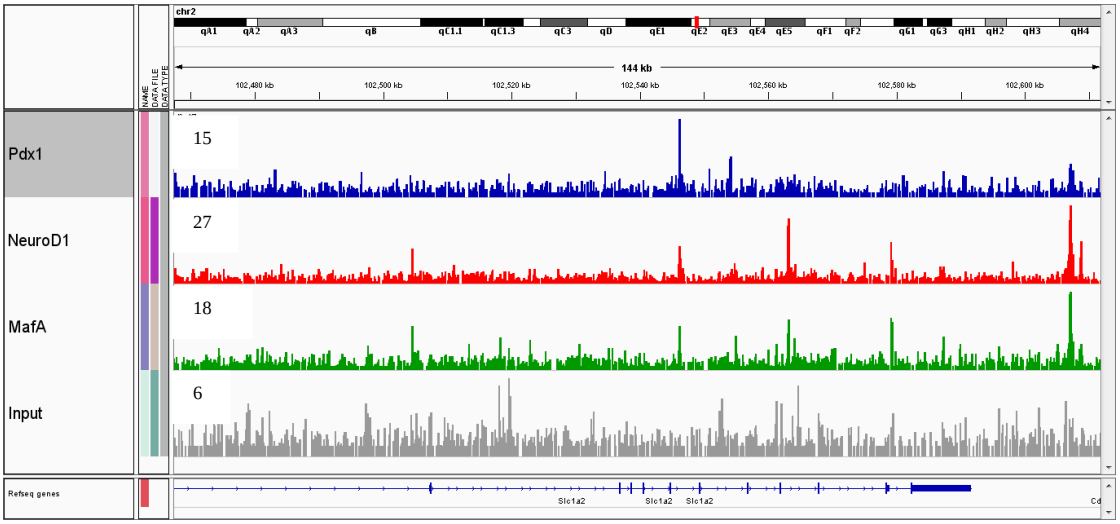
(I) HistH1c



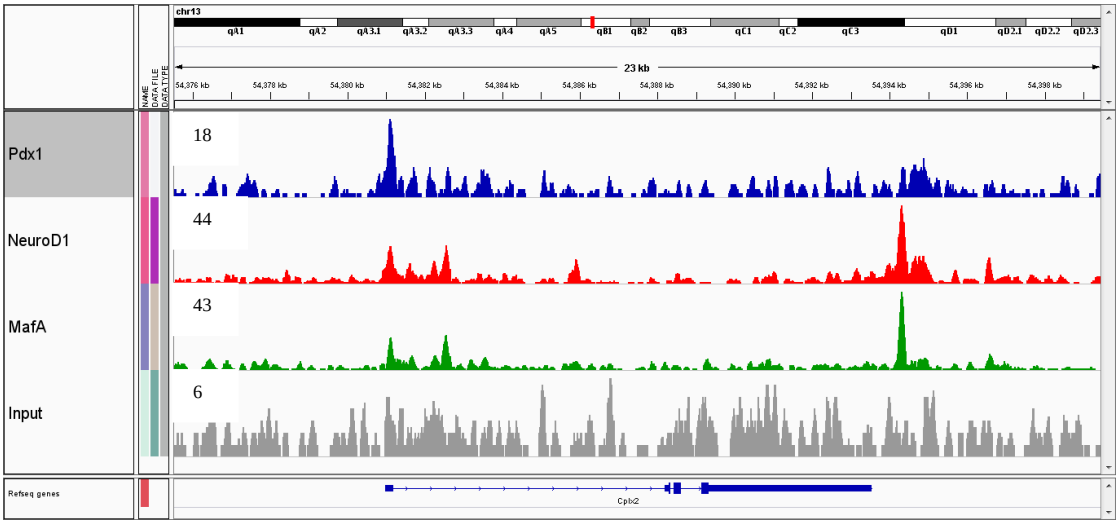
(J) Atp1b1



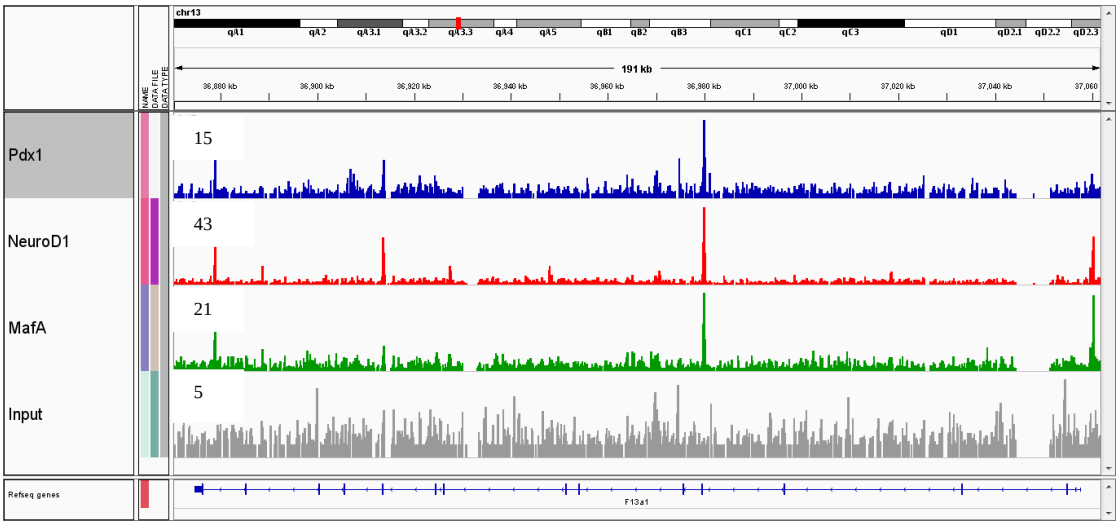
(K) Slc1a2



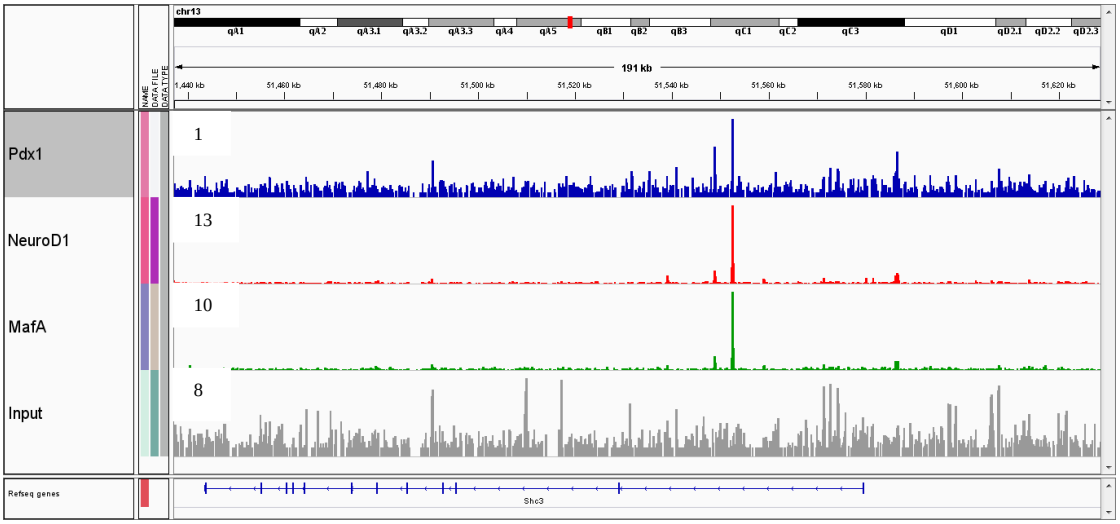
(L) Cplx2



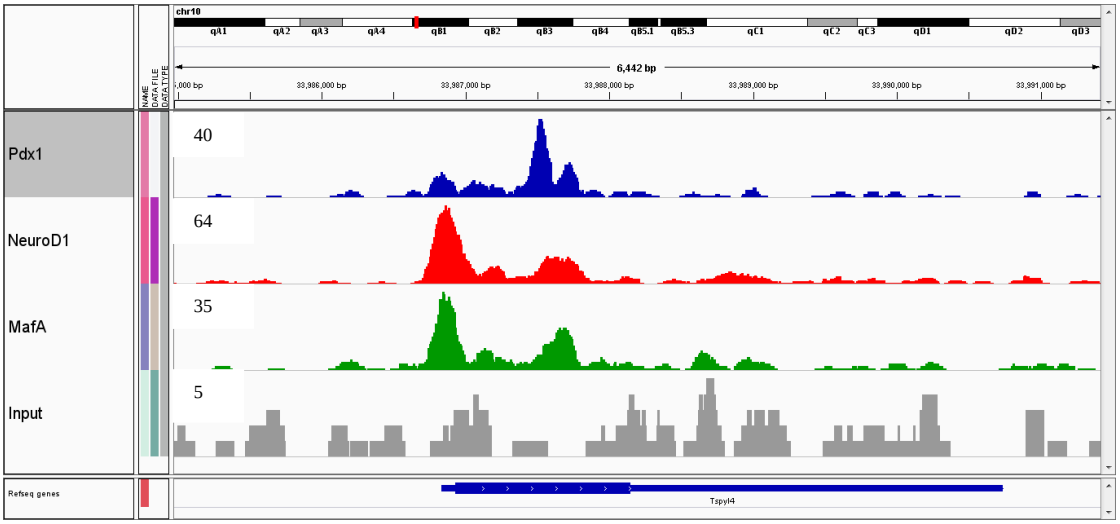
(M) F13a1



(N) Shc3



(O) Tspyl4



(P) Bend7

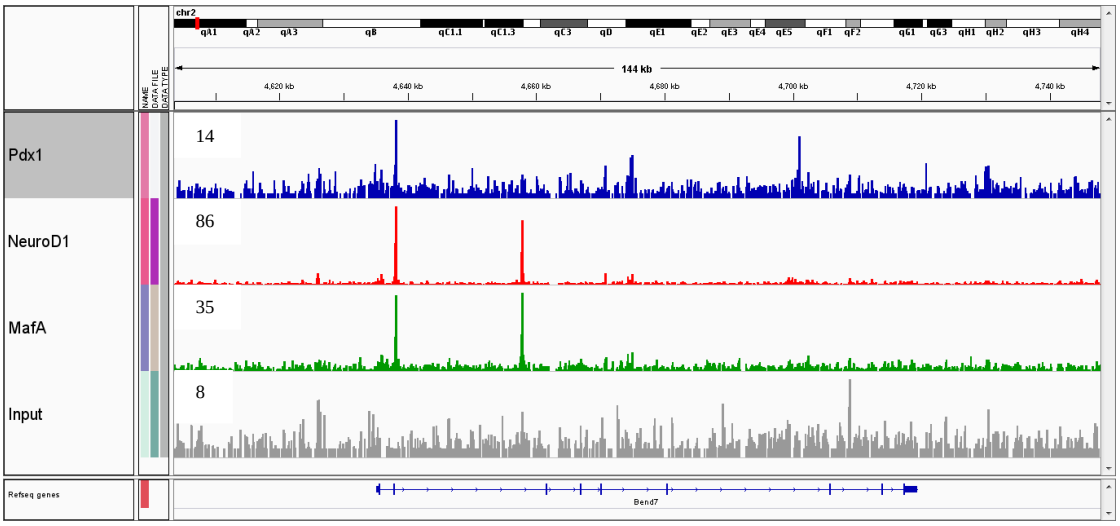


Figure 30. Pdx1, NeuroD1 and MafA binding at 16 of its target genes. Pdx1, NeuroD1 and MafA ChIP-Seq was performed in NIT-1 cells and the binding of all three transcription factors at the (A) *Elfn2* (B) *Exoc4* (C) *Itgb1bp2* (D) *Gcnt2* (E) *Sik2* (F) *Ppp1r3a* (G) *Septin7* (H) *Aldob* (I) *Hist1H1c* (J) *Atp1b1* (K) *Slc1a2* (L) *Cplx2* (M) *F13a1* (N) *Shc3* (O) *Tspyl4* (P) *Bend7* gene promoters is visualized with the IVG[211] software. The input library was included as a control for background signal. Numbers indicate the peak height.

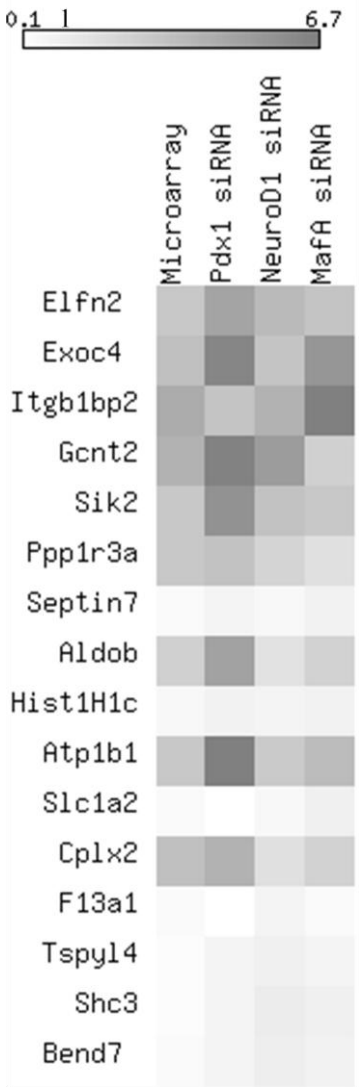


Figure 31. mRNA heatmap of qRT-PCR validation of 16 genes identified to be bound and regulated by Pdx1, NeuroD1 and MafA. NIT-1 cells were transfected with siRNAs against *Pdx1*, *NeuroD1* and *MafA* and RNA was harvested 48hrs post-transfection and transcript levels assayed by RT-PCR. Mean values were plotted relative to NTC (100%) and normalized to *β-actin.* The results are graphically displayed, with the expression levels ranging from 0.1 (white) to 6.7 (gray) and a value of 1 being that of NTC.

4.2.2. TARGET GENES OF PDX1, NEUROD1 AND MAFA REGULATE GSIS IN PANCREATIC BETA-CELLS

As the primary function of the mature β-cell is to secrete insulin in response to glucose, I hypothesized that target genes of Pdx1, NeuroD1 and MafA would be involved in this regulatory process. To address this question, NIT-1 cells were transfected in biological duplicate with siRNAs against the aforementioned 16 genes (Section 4.4.1.). RNA was harvested 48 hours post-transfection and tested for the efficacy of knockdown. As shown in Figure 32, all the target genes showed 50-90% depletion of its gene relative to NTC siRNA. I next went on to see if the ability of the cells to secrete insulin when challenged with glucose was hampered by the introduction of these 16 siRNAs. To that end, GSIS assays were performed on NIT-1 cells transfected with each of the 16 siRNAs in biological triplicates. The amount of insulin secreted when the cells were challenged with glucose vs. basal insulin secretion was quantified with ELISA. siRNAs against *Septin-7*, *Tspy1*, *Cplx2* ($p \leq 0.05$), *Itgb1bp2* and *F13al* ($p \leq 0.1$) resulted in a 20-40% increase in GSIS compared with the NTC siRNA (Figure 33).

The question that subsequently arose was how these 5 genes were mediating their increase in GSIS. To address that, I looked at a panel of β-cell genes involved in various aspects of β-cell function (Table 9) to see if the changes in GSIS were due to changes in expression of some of these genes. Briefly, these genes are all expressed in the pancreatic β-cell and play a role in regulating GSIS. They are *Nkx6.1, Aryl-hydrocarbon receptor nuclear translocator (Arnt)-2, Proprotein convertase subtilisin/kexin type (Pcsk)-1, Ins2, Glut2, Transmembrane (Tmem)-27, Snap25, Vamp2, Stx1a, Calcium channel voltage-dependent P/Q type alpha (Cacna)-1a to c, Styl4, Potassium inwardly-rectifying channel subfamily J member 11 (Kirrel6.2), Syt7, Stxbp1, Uncoupling protein (Ucp)-2, ATPase Ca^{2+} transporting cardiac muscle fast twitch 1 (Serca), and Chromogranin (Chg)-b*. I also looked at genes involved in cell signaling processes: *Protein tyrosine phosphatase receptor type N (Ptprn), Kin of IRRE like (Kirrel)-2, Glutamate decarboxylase (Gad)-1,* and *Iapp*. I also looked at genes that were involved in

protection against apoptosis: *Hnf1α, Igf2, Chloride intracellular channel (Clic)-1* and genes modulating its response to nutrients: *NDRG family member (Ndgr)-2*, and *Gk*.

When Pdx1, NeuroD1 and MafA are depleted from NIT-1 cells, there is an increase in mRNA levels of *Itgb1bp2* and *Cplx2* (Figure 31). Upon knockdown of *Itgb1bp2* (Figure 34) and *Cplx2* (Figure 35), there is a 20% increase ($p \leq 0.1$) and 30% increase in GSIS ($p \leq 0.05$) respectively, with a simultaneous ≥1.5 fold increase in levels of *Nkx6.1* and *Glut2* relative to NTC siRNA. There was also a ≥1.5 fold increase in mRNA levels of genes involved in the insulin secretory process upon *Itgb1bp2* knockdown, e.g. *Snap25*, *Cacna1b*, *Kirrel6.2*, *Syt7*, *Stxbp1*, *Serca* ($p \leq 0.05$). On top of those genes that increased in mRNA expression, in samples treated with siRNAs against *Cplx2*, there were more insulin secretory genes that showed a ≥1.5 fold increase in expression, e.g. *Vamp2*, Syt1a, *Cacna1a* and *1c*, and *Chgb* ($p \leq 0.05$). This suggests that Pdx1, NeuroD1 and MafA bind directly to and repress *Itgb1bp2* and *Cplx2* to stimulate the GSIS response in β-cells via increasing the transcription of important β-cell genes, including those involved in the GSIS secretory response. There also appears to be a slight anti-apoptotic effect mediated by Itgb1bp2 and Cplx2, in that decreasing both of them led to a decrease in Clic1. Clic1 is a target gene of NeuroD1 and the increase in Clic1 leads to hyperglycemia-induced apoptosis[113].

When Pdx1, NeuroD1 and MafA were depleted from cells, there was a decrease in mRNA levels of *Tspyl1*, *F13a1* and *Septin7* (Figure 31). Upon knockdown of *Tspyl1*, *F13a1* and *Sept7*, there was a 30-40% increase in GSIS (*Tspyl1* and *Septin7*: $p \leq 0.05$, F13a1: $p \leq 0.1$), with a concomitant ≥1.5 fold increase in mRNA levels of *IAPP*, *Ndrg2*, *Glut2* and *Gad1* ($p \leq 0.05$) (Figures 36–38). There was also a ≥1.5 fold increase in expression of genes involved in the insulin secretory response, e.g. *Snap25*, *Vamp2*, *Cacna1b*, *Kirrel6.2*, *Stxbp1*, *Serca* and *Chgb* ($p \leq 0.05$). This suggests that Pdx1, NeuroD1 and MafA bind to and increase the expression of *Tspyl1*, *F13a1*, and *Sept7*, possibly resulting in an indirect reduction of GSIS as depletion of these genes have an increase in GSIS.

Figure 32. mRNA heatmap of NIT-1 cells transfected with 16 siRNAs. NIT-1 cells were transfected in duplicate with 16 siRNAs against genes I identified to be Pdx1, NeuroD1 and MafA targets. RNA was harvested 48hrs post-transfection and transcript levels assayed by RT-PCR. Mean values were plotted relative to NTC (100%) and normalized to *β-actin*. The results are graphically displayed, with the expression levels ranging from 0.11 (white) to 5 (gray) and a value of 1 being that of NTC.

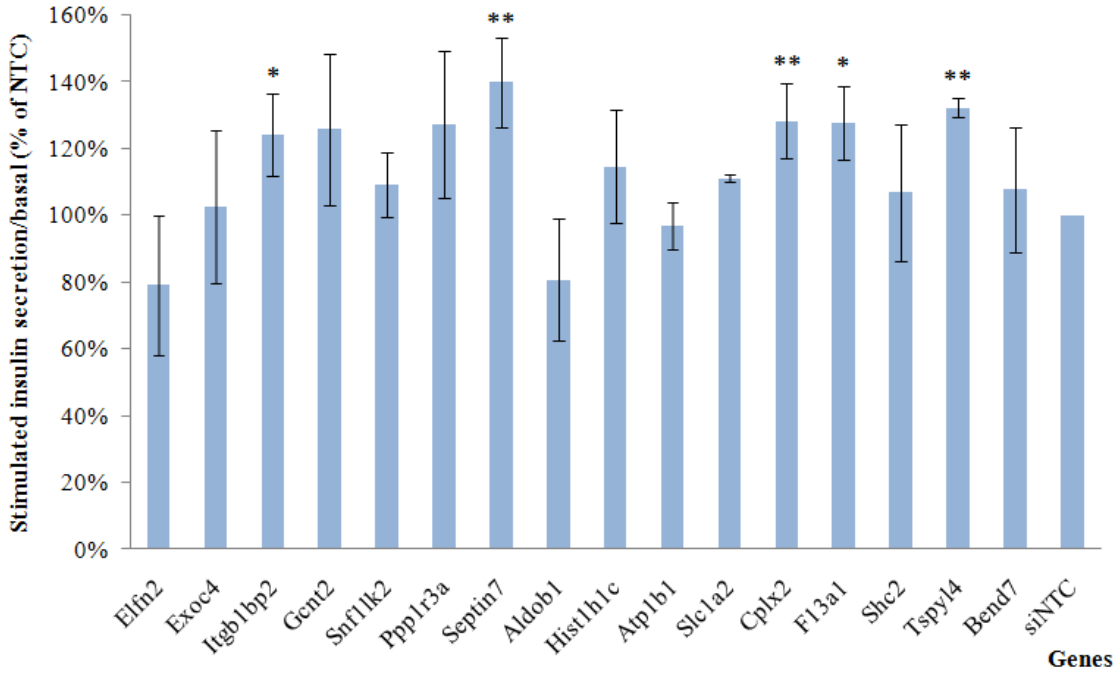


Figure 33. GSIS of NIT-1 cells transfected with siRNAs against 16 genes triply bound and regulated by Pdx1, NeuroD1 and MafA. NIT-1 cells were transfected in triplicate with siRNAs against 16 genes identified to be Pdx1, NeuroD1 and MafA targets, and GSIS assays were performed. The cells were starved for 2 hours at 37°C before incubation with 1xKRH for 1 hour (basal secetion) before incubation with 1XKRH with 20mM glucose for 1 hour (stimulated secretion). Released insulin in the supernatent was quantified with ELISA in technical triplicates and the values are means ± SE (n=3). **$p \leq 0.05$, *$p \leq 0.1$. siRNAs against Septin7, Cplx2, Tspyl4, Itgb1bp2 and F13a1 resulted in significant decreases in GSIS.

Gene	Function
Nkx6.1	Expressed in developing and mature pancreas and brain. Functions downstream of Nkx2.2 in pancreas formation[73]. Endodermal expression of genes depend on Pdx1[240]. Suppresses Gcg expression and regulates GSIS in β-cells.[75]
HNF1α	Plays an important role in pancreatic β-cell differentiation and survival[241]. Heterozygous Hnf1α mutations cause pancreatic islet β-cell dysfunction and MODY3[242]. The expression of Glut2 is decreased in pancreatic islets upon introduction of a dominant negative form of Hnf1α[243].
Arnt2	Modulation of Arnt2 in pancreatic β-cells affects GSIS[244]
Igf2	Increased pancreatic Igf2 in undernourished rats may contribute to the partial suppression of the developmental wave of β-cell apoptosis .
Iapp	37-residue peptide hormone secreted by pancreatic β-cells at the same time as insulin, in a roughly 1:100 amylin:insulin ratio. Major component of the amyloid fibrils found in islets of T2DM patients[245].
Ptprn	Member of the protein tyrosine phosphatase family, which are signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation[246]. Auto-antigen in T1DM, intrinsic membrane protein of neurosecretory vesicles [247].
Ndrg2	Highly expressed in pancreatic β-cells and is involved in the Akt-mediated protection of β-cells against lipotoxicity[248].
Pcsk1	Pcsk1 and 2 differentially cleave pro-opiomelanocortin and they act together to process proinsulin and proglucagon in pancreatic islets.[249]
Kirrel2	A novel immunoglobulin superfamily gene expressed primarily in β-cells of the pancreas[250].
Insulin2	Profound glucose intolerance resulted from reduced insulin secretion accompanied by abnormal distension of the ER lumen, defective trafficking of proinsulin and a reduced number of insulin granules in β-ells [251].
Clic1	A target gene of NeuroD1 in the pancreas, it is expressed in the developing and adult islet and is involved in GSIS. When glucose is elevated, the expression of Clic1 sensitizes β-cells to high glucose stress, which in turn, causes ER stress. As a result, expression of Clic1 increased hyperglycemia-induced apoptosis. In addition, the expression of Clic1 under high glucose conditions decreased the expression of genes important for β-cell function, such as Gk, Pdx1, and Insulin[113].

Gene	Function
Glut2	A glucose transporter essential for GSIS, thereby controlling blood glucose homeostasis response to dietary intake. Its expression is strongly reduced in glucose-unresponsive islets from different animal models of diabetes[252] .
Tmem27	A target of Hnf1, it controls insulin secretion by SNARE formation. Expression of Tmem27 is reduced in Hnf1$\alpha^{-/-}$ mice and is increased in islets of mouse models with hypertrophy of the endocrine pancreas[253].
Gad1	This gene encodes one of several forms of glutamic acid decarboxylase and is identified as an autoantigen and an autoreactive T cell target in IDDM [254].
Gk	Catalyzes the initial step in utilization of glucose by the β-cell and liver at physiological glucose concentrations. Gk has a high K_m for glucose, and so it is effective only when glucose is abundant. The role of Gk is to provide glucose-6-phosphate for the synthesis of glycogen. Pancreatic Gk plays an important role in modulating insulin secretion[255].
Snap25	Component of the SNARE complex, which is proposed to account for the specificity of membrane fusion and to directly execute fusion by forming a tight complex that brings synaptic vesicles and plasma membranes together[256].
Vamp2	Synaptobrevins/Vamps, syntaxins, and Snap25 are the main components of a protein complex involved in the docking and/or fusion of synaptic vesicles with the presynaptic membrane. Vamp2 is thought to participate in neurotransmitter release at a step between docking and fusion[256].
Stx1a	Syntaxins are nervous system-specific proteins implicated in the docking of synaptic vesicles with the presynaptic plasma membrane. Syntaxins bind synaptogamin in a calcium-dependent fashion and interact with voltage dependent calcium and potassium channels. Syntaxin-1A is a key protein in ion channel regulation and synaptic exocytosis.[257]
Cacna1a,b,c	Voltage-gated calcium channels (CaV) are present in the membrane of most excitable cells and mediate calcium influx in response to depolarisation. They regulate intracellular processes such as contraction, secretion, neuro-transmission and gene expression[258].
Sytl4	This gene encodes a member of the synaptotagmin -like protein family. The encoded protein binds specific small Rab GTPases and is involved in intracellular membrane trafficking. This protein binds Rab27 and is involved in inhibiting insulin dense core vesicle exocytosis[259].

Gene	Function
Kirrel6.2	K_{ir}6.2 is a major subunit of the ATP-sensitive K^+ channel, an inward-rectifier potassium ion channel. The gene encoding the channel is KCNJ11and mutations in this gene are associated with congenital hyperinsulinism[260].
Syt7	Synaptotagmins, such as SYT7, are brain-specific calcium-dependent phospholipid-binding proteins that play a role in synaptic exocytosis and neurotransmitter release. It is required for the maintenance of systemic glucose tolerance and GSIS[261].
Stxbp1	Regulation of synaptic vesicle docking and fusion, possibly through interaction with GTP-binding proteins. Essential for neurotransmission and binds syntaxin, a component of the synaptic vesicle fusion machinery probably in a 1:1 ratio[256].
Ucp2	Hyperglycemia-induced mitochondrial superoxide production activates Ucp2, which decreases the ATP/ADP ratio and thus reduces the insulin secretory response[262].
Serca	SERCA, or sarco/endoplasmic reticulum Ca^{2+}-ATPase, is a calcium ATPase-type P-ATPase. Insulin regulation of β-Cell function involves a feedback loop on Serca gene expression, Ca^{2+} homeostasis, and Insulin expression and secretion.[263]
Chgb	This gene encodes a tyrosine-sulfated secretory protein abundant in peptidergic endocrine cells and neurons. This protein may serve as a precursor for regulatory peptides. It is not required for normal insulin granule biogenesis or maintenance in vivo, but is essential for adequate secretion of islet hormones[264].

Table 9. β-cell genes involved in various aspects of β-cell function.

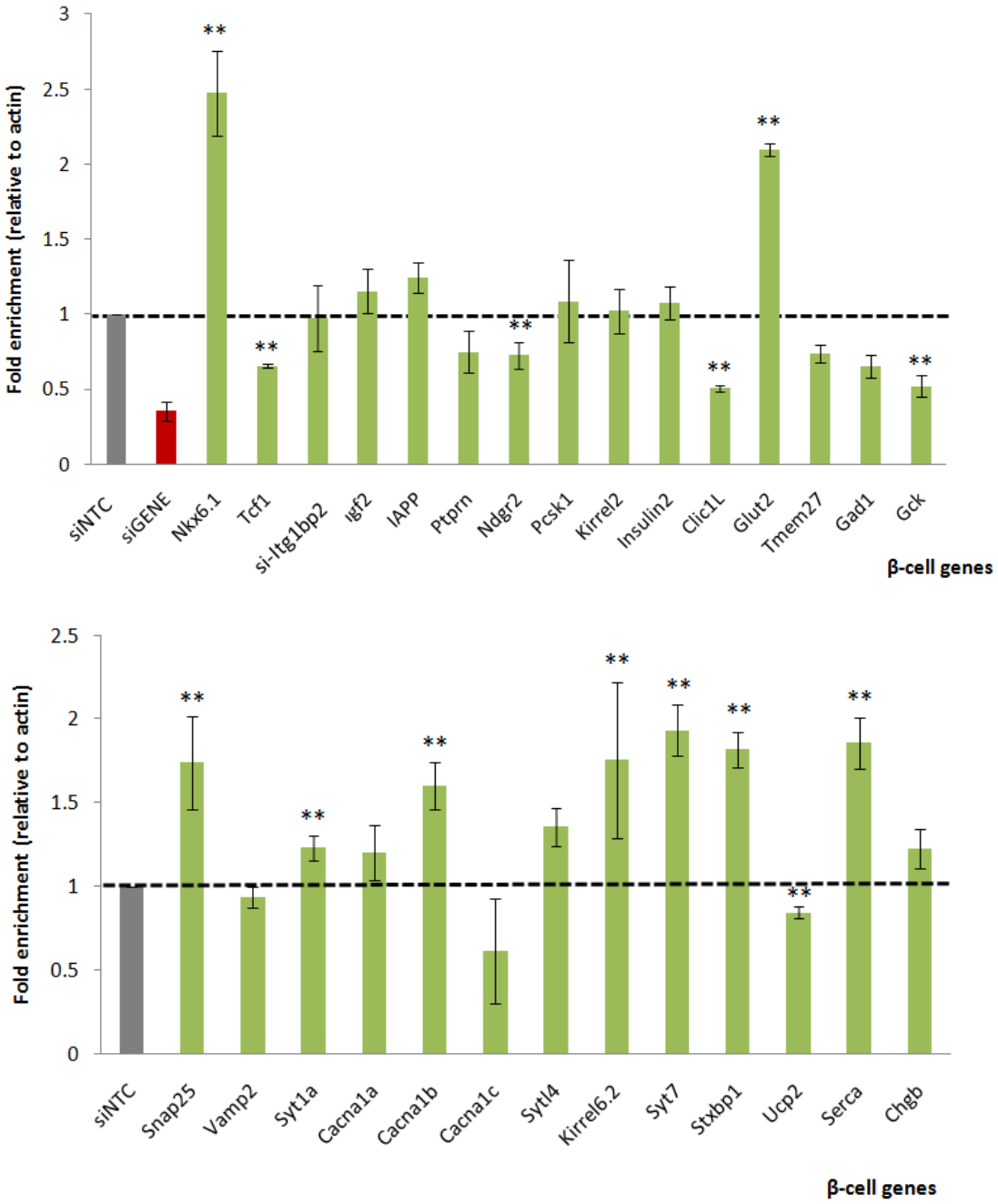


Figure 34. *Itgb1bp2*-depletion in NIT-1 cells leads to an increase in expression of key β-cell genes. NIT-1 cells were transfected with siRNAs against *Itgb1bp2* and RNA was harvested 48hr post-transfection. Transcript levels of important β-cell genes were analyzed with qRT-PCR and normalized against *ß-actin*. The transfections were performed in biological duplicates and the RT-PCRs in technical duplicates, and the average of the normalized ratio of target gene/ß-actin was calculated and presented with standard deviation. **$p \leq 0.05$.

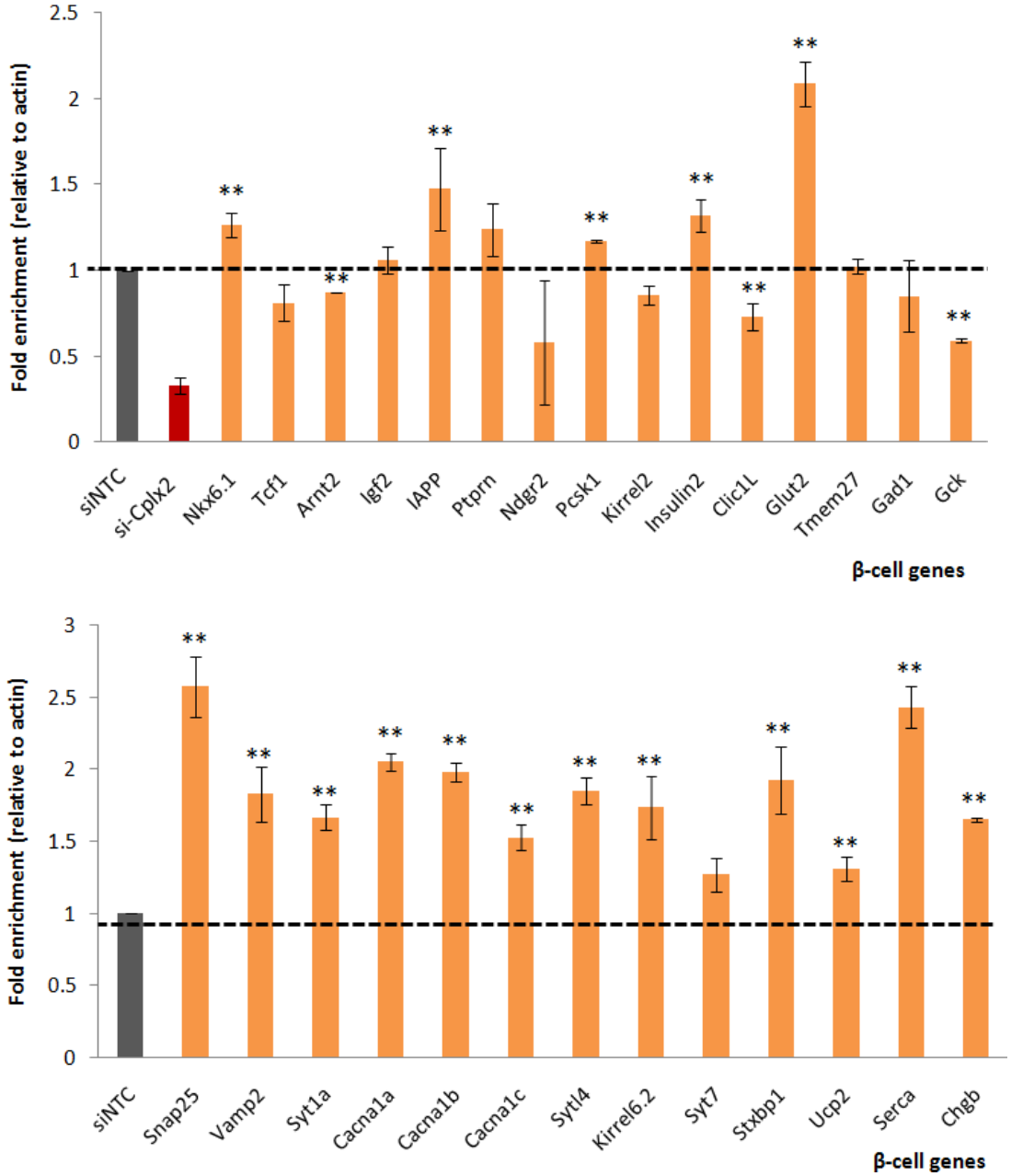


Figure 35. *Cplx2*-depletion in NIT-1 cells leads to an increase in expression of key β-cell genes. NIT-1 cells were transfected with siRNAs against *Cplx2* and RNA was harvested 48hr post-transfection. Transcript levels of important β-cell genes were analyzed with qRT-PCR and normalized against *ß-actin*. The transfections were performed in biological duplicates and the RT-PCRs in technical duplicates, and the average of the normalized ratio of target gene/ß-actin was calculated and presented with standard deviation. **$p \leq 0.05$.

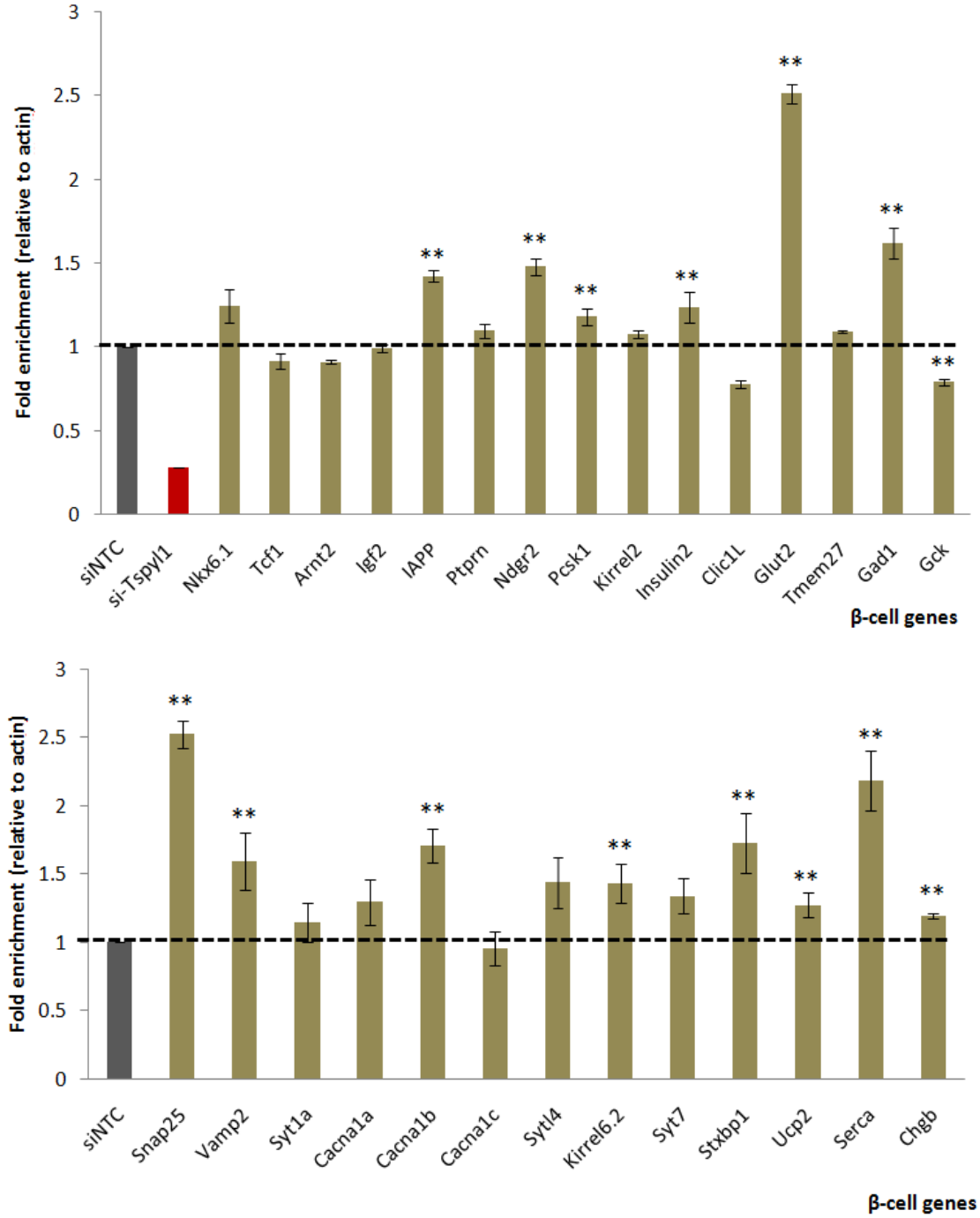


Figure 36. *Tspyl1*-depletion in NIT-1 cells leads to an increase in expression of key β-cell genes. NIT-1 cells were transfected with siRNAs against *Tspyl1* and RNA was harvested 48hr post-transfection. Transcript levels of important β-cell genes were analyzed with qRT-PCR and normalized against *ß-actin*. The transfections were performed in biological duplicates and the RT-PCRs in technical duplicates, and the average of the normalized ratio of target gene/ß-actin was calculated and presented with standard deviation. **$p \leq 0.05$.

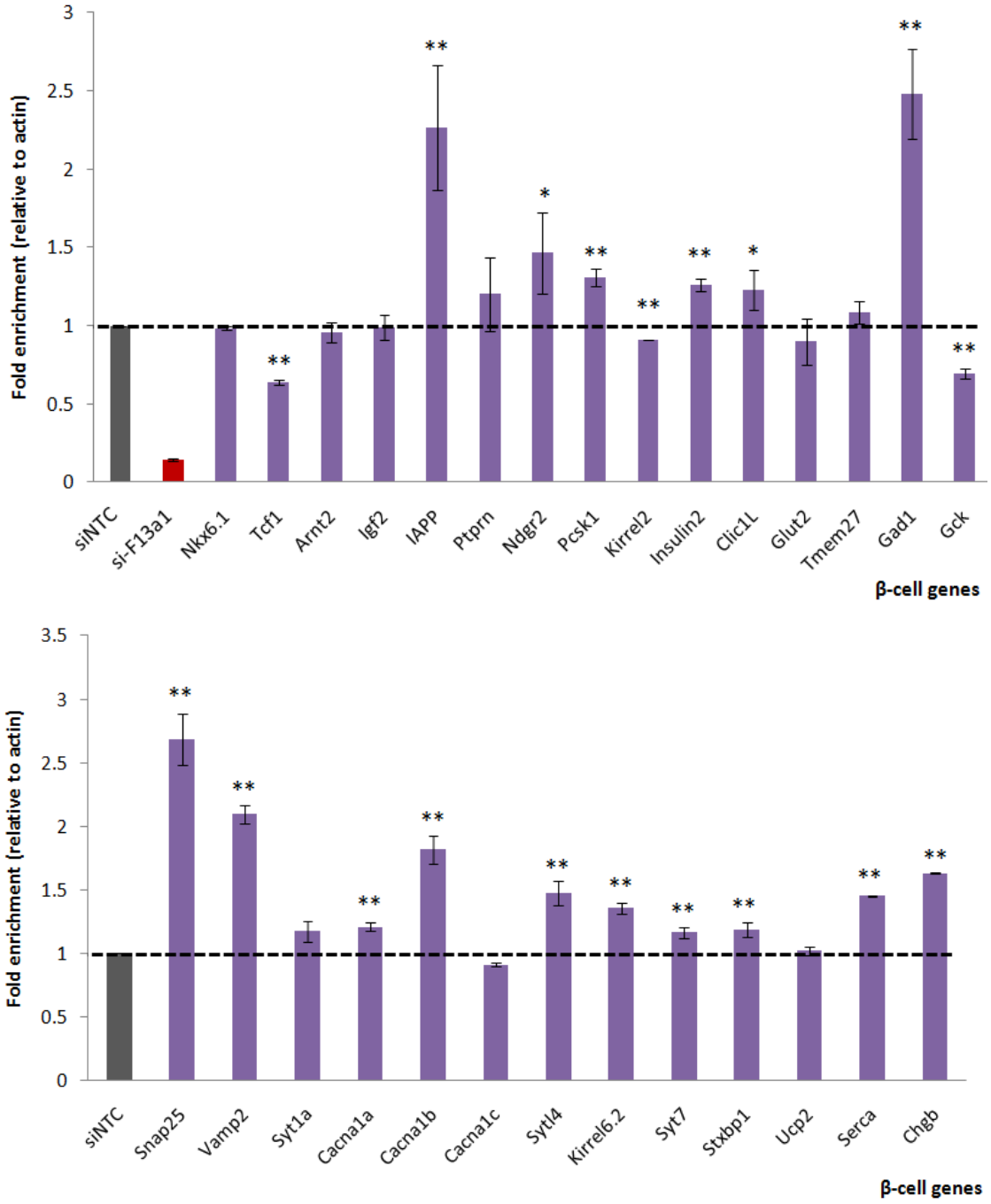


Figure 37. *F13a1*-depletion in NIT-1 cells leads to an increase in expression of key β-cell genes. NIT-1 cells were transfected with siRNAs against *F13a1* and RNA was harvested 48hr post-transfection. Transcript levels of important β-cell genes were analyzed with qRT-PCR and normalized against *ß-actin*. The transfections were performed in biological duplicates and the RT-PCRs in technical duplicates, and the average of the normalized ratio of target gene/ß-actin was calculated and presented with standard deviation. **$p \leq 0.05$, * $p \leq 0.1$.

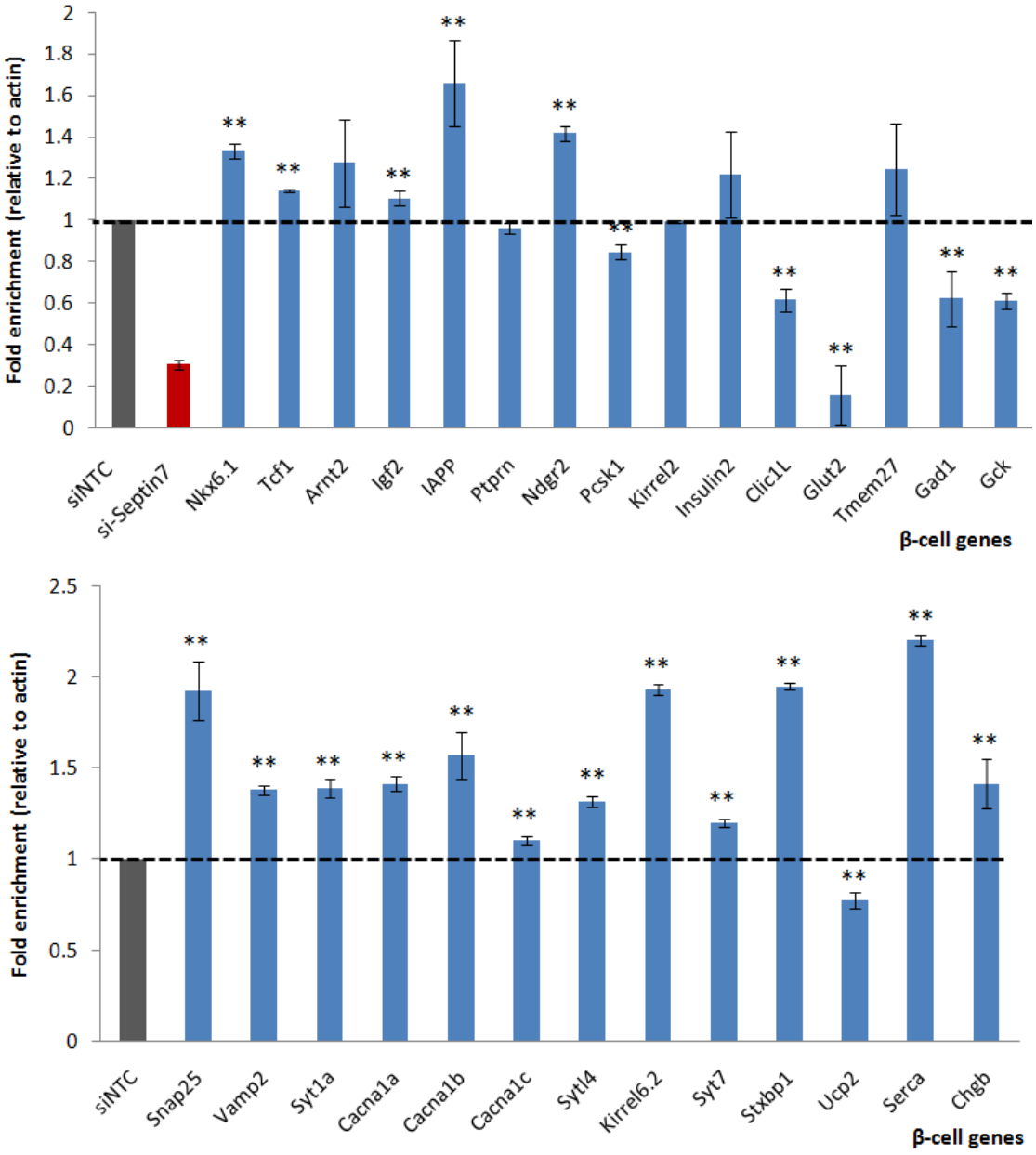


Figure 38. *Septin7*-depletion in NIT-1 cells leads to an increase in expression of key β-cell genes. NIT-1 cells were transfected with siRNAs against *Septin7* and RNA was harvested 48hr post-transfection. Transcript levels of important β-cell genes were analyzed with qRT-PCR and normalized against *ß-actin*. The transfections were performed in biological duplicates and the RT-PCRs in technical duplicates, and the average of the normalized ratio of target gene/ß-actin was calculated and presented with standard deviation. **$p \leq 0.05$.

4.3. DISCUSSION

Pancreatic β-cells are able to sense glucose and other nutrient secretagogues to regulate insulin exocytosis, thereby maintaining glucose homeostasis. This systems biology of insulin secretion controls translation of metabolic signals into intracellular messengers recognized by the exocytotic machinery. Pdx1, NeuroD1 and MafA are key transcriptional regulators of the mature β-cell phenotype, with their target genes having a range of regulatory effects in many aspects, such as cellular metabolism, developmental processes, and cellular activities such as regulation of cell cycle, cell death and proliferation (Section 3). These genes bound by these three transcription factors are also associated with genetic diseases, neurological and metabolic disorders and gastrointestinal disease. Because of their importance in regulating the mature β-cell phenotype, I hypothesized that the genes bound to and regulated by all three transcription factors would play an important role in the main function of the mature β-cell, which is to secrete insulin in response to glucose.

4.3.1. ITGB1BP2 AND CPLX2 POSITIVELY REGULATE GSIS

When Pdx1, NeuroD1 and MafA were depleted from NIT-1 cells, there was an increase in mRNA levels of *Itgb1bp2* (Figure 31). Upon knockdown of *Itg1bp2*, there was a 20% increase ($p \leq 0.1$) in GSIS, with a ≥1.5 fold increase in expression of many key β-cell genes (Figure 34). This suggests that Pdx1, NeuroD1 and MafA bind to and repress *Itgb1bp2* to increase the GSIS response in β-cells via increasing the transcription of important β-cell genes, including those involved in the GSIS secretory response. There was also a slight anti-apoptotic effect mediated by Itgb1bp2, in that its depletion led to a decrease in expression of *Clic1*. *Clic1* is a target gene of NeuroD1 and the increase in *Clic1* in hyperglycaemic conditions sensitizes mouse β-cells to high glucose stress, which causes ER stress and apoptosis[113]. The expression of *Clic1* under high glucose conditions also decreased the expression of *Pdx1*, *Gk* and *Insulin*[113]. Therefore, perhaps Pdx1, NeuroD1 and MafA indirectly reduces the expression of *Clic1* in normal β-cells via Itgb1bp2, to protect the cells from stress and apoptosis.

The only known role of Itgb1bp2 is its involvement during maturation and/or organization of muscles cells[222]. It interacts with the β-integrin cytoplasmic domain and appears to be regulated by intracellular signals affecting Ca^{2+} concentration. The ability of pancreatic β-cells to respond to glucose has long been known to depend on extracellular Ca^{2+} levels[265], and release of intracellular Ca^{2+} stores (by the influx of extracellular Ca^{2+} or by other external signals) is also thought to be involved in regulating insulin secretion in order to enhance or prolong insulin secretion instead of directly triggering it[266]. It is not inconceivable that similar mechanisms could be similarly involved here, with Pdx1, NeuroD1 and MafA responding to extracellular Ca^{2+} signals via Itgb1bp2, in order to indirectly influence the β-cell's GSIS response.

A 30% increase in the GSIS response ($p \leq 0.05$) was seen when *Cplx2*, another target of Pdx1, NeuroD1 and MafA, was depleted from β-cells (Figure 35). This increase in GSIS also involved the ≥1.5 fold up-regulation of important β-cell genes, especially those involved in mediating the GSIS response. Similar to the effects of *Itgb1bp2*-depletion, there was also a decrease in the anti-apoptotic gene *Clic1*, suggesting a minor role in Pdx1, NeuroD1 and MafA indirectly reducing the expression of *Clic1* in normal β-cells via Cplx2 to protect against apoptosis.

Cplx is a small protein that binds rapidly with high affinity to the SNARE complex, which is a multistage process central to synaptic transmission and hormone release; therefore Cplx has a regulatory role in fast exocytosis[267]. There are 4 *Cplx* isoforms in vertebrates, all of them highly enriched in the nervous system, with some present in other tissues. The function of Cplxs in regulated exocytosis is controversial, with several early studies suggesting a negative role in the release of neurotransmitters[268] such as acetylcholine[269], and spontaneous neurotransmitter release upon knockout of *Cplx* was observed[270]. Other studies indicate that Cplxs plays a positive role in exocytosis, with adrenal chromaffin cells from *Cplx2* knockout mice exhibiting markedly diminished releasable vesicle pools[237], and insulin secretion was

diminished in pancreatic β-cells both over- and under-expressing *Cplx1*[271]. *Cplx2* is expressed at low levels in different tissues, and in neuronal cells has a function similar to Cplx1 in the control of SNARE-mediated exocytosis[272]. The mouse MIN6 and rat INS1 β-cell line do not expresses *Cplx2*, but the mouse β–TC6 cell line expresses low levels of *Cplx2*, unlike all the pancreatic β-cell lines which express *Cplx1* at high levels[271], suggesting that the expression of this gene is regulated in a β-cell line-dependent manner (possibly due to clonal selection during generation of cell lines). *Cplx1* and *Cplx2* are encoded by two different genes and are probably differentially regulated in β-cells, as demonstrated by the differential expression levels and cellular localizations of Cplx1 and Cplx2 in the rat forebrain and hippocampus[273].

I have shown that *Cplx2* is repressed by Pdx1, NeuroD1 and MafA, and upon *Cplx2* knockdown in the NIT-1 β-cell line, there is a 30% increase in GSIS, together with the increase in expression of key β-cell genes, including those involved in the insulin secretory response (Figure 35). This indicates Cplx2 plays a repressive role in the GSIS response in pancreatic β-cells, unlike Cplx1. It also appears it is necessary for Pdx1, NeuroD1 and MafA to repress *Cplx2* in β-cells in order to maintain a robust GSIS response.

4.3.2. TSPYL1, F13A1 AND SEPTIN7 NEGATIVELY REGULATE GSIS

F13a1 has a well-established role in blood clotting, it encodes the coagulation factor XIII A subunit, which is the last zymogen to become activated in the blood coagulation cascade[238]. It is expressed in the pancreas but has no known function there. *Tspyl1* is expressed in the pancreas and has no known function at all. *Septin7* is an E2F1/2 responsive gene and is typically regulated at G2/M in the cell cycle[274]. It is a filament-forming cytoskeletal GTPase and is required for normal organization of the actin cytoskeleton[229] and cytokinesis[230].

When Pdx1, NeuroD1 and MafA were depleted from cells, there was a decrease in mRNA levels of *Tspyl1*, *F13a1* and *Septin7* (Figure 31). Upon knockdown of *Tspyl1*, *F13a1* and

Sept7, there was a 30-40% increase in GSIS (Tspyl1 and Septin7: $p \leq 0.05$, F13a1: $p \leq 0.1$), with a concomitant ≥1.5 fold increase in key β-cell genes (Figure 36-38). This suggests that Pdx1, NeuroD1 and MafA bind to and increase the expression of *Tspyl1*, *F13a1*, and *Sept7*, possibly resulting in an indirect reduction of GSIS as depletion of these genes lead to an increase in GSIS.

This is perhaps counter-intuitive to the positive effect on GSIS of Pdx1, NeuroD1 and MafA via Itgb1bp2 and Cplx2, as mentioned in Section 4.3.1. However, it is not surprising that there are genes negatively involved in regulating GSIS. Activation of the α subunit of the heterotrimeric G protein $G\alpha_z$ inhibits GSIS *in vitro* and *in vivo*, with increased insulin secretion observed in $G\alpha_z^{-/-}$ islets due to the relief of a tonic inhibition of adenylyl cyclase (cAMP production was significantly increased in $G\alpha_z^{-/-}$ islets in the absence of exogenous stimulation)[275]. Activation of β-cell M_3 muscarinic receptors (M3R) enhances GSIS via stimulation of G proteins of the G_q family, which are under the inhibitory control of regulators of G protein signaling (RGS) proteins. RGS4 was identified as a potent negative regulator of M3R function in pancreatic β-cells. The pancreatic β-cell needs to respond in the appropriate temporal manner to external signals, it should not be secreting insulin indiscriminately in an excessive manner, nor should it be inhibited in its ability to respond, given the proper stimulus. Taken together, these results clearly demonstrates the complexity of regulation in GSIS, and highlights the dual roles of Pdx1, NeuroD1 and MafA in maintaining the β-cell phenotype.

4.4. SUMMARY

A fundamental property of pancreatic β-cells is their capacity to regulate the release of insulin and other hormones in direct response to changes in extracellular glucose concentration. In large part, this ability defines β-cell function since insulin is produced and released in the body only from these specialized cells. Insulin is crucial to regulating blood glucose, and reduced insulin secretion is a key feature of both T1DM and T2DM, and therefore, GSIS is a well-accepted measure of β-cell function.

I have shown that Pdx1, NeuroD1 and MafA repress *Itgb1bp2* and *Cplx2* in order to achieve an enhanced GSIS response. Pdx1, NeuroD1 and MafA activate *Tspyl1*, *F13a1* and *Septin7* to reduce GSIS (Figure 39). This additional knowledge emphasizes the exquisite level of regulation required for the appropriate response of β-cells to a glucose stimulus.

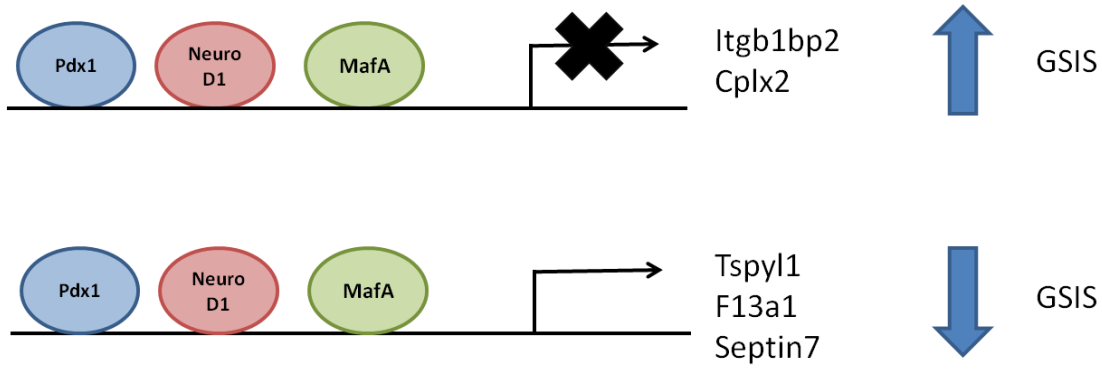


Figure 39. Schematic diagram of how Pdx1, NeuroD1 and MafA interact with their target genes to modulate the insulin secretory response of β-cells to a glucose stimulus.

CHAPTER 5:

EPIGENETICS & THE DYSFUNCTIONAL BETA-CELL

5.1. INTRODUCTION

T2DM develops as a result of insulin resistance coupled with impaired β-cell function, and is closely associated with increased plasma FFA concentrations and elevated glucose concentrations[169]. Continuous over-stimulation of the β-cell by glucose leads to depletion of insulin stores, excessively high levels of ROS, worsening of hyperglycemia, and finally deterioration of the β-cell (glucotoxicity)[276]. FFAs enhance the β-cell response on acute application but long-term exposure exerts lipotoxic effects and leads to blunted GSIS and decreased cell viability[277]. The combination of FFAs and high glucose has been suggested to be particularly deleterious, and the term "glucolipotoxicity" has been coined to describe this phenomenon[169].

Pdx1 associates with HDAC1 and -2 in low glucose conditions (1-3mM) to repress INS gene expression, and recruits the co-activator HAT p300 to the *Insulin* promoter only in high glucose conditions (10-30mM)[161]. The HMT Set7/9 is recruited by Pdx1 to the *Insulin* and *Glut2* genes, leading to H3K4me2 and recruitment of RNA polymerase II[162]. Pdx1 modulates histone H4 acetylation to directly activate *Insulin* gene expression in β-cells; exogenous *Pdx1* expression in α-cells also leads to *Insulin* expression via H4 acetylation[164]. Expression of *Pdx1* is permanently reduced in IUGR rats when the *Pdx1* gene undergoes epigenetic modifications throughout development leading to its silencing[167]. *Pdx1* and *MafA* expression and activity are decreased in DIO animals and in cell culture conditions mimicking diabetes, leading to a decrease in insulin biosynthesis and secretion[169]. A mutation in p300 removes binding to *NeuroD1* and destroyed the ability of p300 to activate *Insulin* E-box directed transcription in β-cells, leading to MODY6[166]. Therefore, the association of many β-cell transcriptional regulators with HATs, HMTs and HDACs provide correlative evidence for histone modification in regulating β-cell function or growth.

The mono-methylations of H3K27, H3K9, H4K20, H3K79, and H2BK5 are linked to gene activation, while tri-methylations of H3K27, H3K9, and H3K79 are linked to repression[136]. The apparently opposite modifications H3K4me3 and H3K27me3 co-localize in ‘‘bivalent domains’’ in embryonic stem cells[137], and these have been suggested to function in the differentiation of these cells, as these “bivalent domains” have been found in developmental regulators. Recent evidence has confirmed the presence of these domains in mouse embryonic fibroblasts, neural progenitors[138] and human T cells[139], proving that these “bivalent domains” are not an ESC-specific phenomenon.

The human INS gene has recently been found to be part of a large open chromatin domain specific in islets which is marked by histone acetylation and H3K4 di-methylation marks distributed over the entire coding region and an 80kb region around it[176]. In another recent study that looked at open regions of chromatin in human pancreatic islets[177], the authors identified islet-specific COREs, of which many genes encoding proteins with known function in islets (such as PDX1, PAX6, GLP1R, SLC30A8 and CPE), belong[178]. In another new study, open chromatin regions, CTCF-binding sites, H3K4me3, H3K4me1, and H3K79me2 were profiled across the entire genome in human islets[179]. Integrated analysis of these large-scale data sets identified 18000 putative TSSs, of which 30% were previously unannotated by RefSeq and at least several hundred of these are islet-active TSSs, including those for major islet miRNAs previously implicated in the control of glucose homeostasis[180]. In summary, there is a growing body of current literature suggesting a role for epigenetic interactions in the complex interplay between gene regulation and the environment.

Therefore, my hypothesis is that is that as β-cell dysfunction progressively develops, modulation of the chromatin environment adversely affects the expression and activity of Pdx1, MafA and NeuroD1 and other key β-cell transcriptional regulators.

Questions I will address in this chapter are:

1) Where are the regions of active and repressive chromatin in normal and dysfunctional β-cells?

2) Are there differences in open and closed regions of chromatin in normal and β-cells, and how do these change upon β-cell becoming dysfunction?

3) How does the global Pdx1, NeuroD1 and MafA occupancy correlate with histone marks in normal β-cells, and do these patterns change upon β-cell dysfunction?

4) Are there bivalent domains in normal β-cells? Do these domains change upon β-cell dysfunction?

To answer these questions, I have profiled H3K4me3 (an active mark of chromatin), H3K27me3 (a repressive mark of chromatin) in normal β-cells to identify regions of chromatin that are actively transcribed and those that are not, and compared that with the same histone marks identified in β-cells that have a decrease in GSIS after treatment with high glucose and fatty acids, i.e. dysfunctional β-cells. I have also used the FAIRE technique[188] to identify regions of open and closed chromatin in normal and dysfunctional β-cells. FAIRE performed in human cells strongly enriches DNA coincident with the location of DNaseI hypersensitive sites, TSS, active promoters, and also displays cell-type–specific patterns[188].

5.2. RESULTS

5.2.1. GENERATION OF DYSFUNCTIONAL BETA-CELLS

The main function of the β-cell is to secrete insulin in response to a glucose stimulus. Response to an environmental stimulus is not only a metabolic one, the physiological long term adaptation of pancreatic β-cells is driven by stimuli such as glucose and incretin hormones acting via cAMP (e.g. Glp-1) and involves regulated gene expression. Several rapidly inducible immediate-early genes (IEGs) have been identified in β-cells. Many of these IEGs code for transcription factors and have the potential to control the transcription of downstream target genes likely involved in long term cellular adaptation[278].

Exposure of isolated islets and insulin-secreting cells to elevated levels of fatty acids *in vitro* impairs *Insulin* gene expression when glucose concentrations are concomitantly elevated[57]. Palmitate and oleate inhibit insulin secretion, but only palmitate impairs *Insulin* gene expression, with palmitate impairment of the *Insulin* gene appearing to be mediated by direct inhibition of Insulin promoter activity[279]. Palmitate inhibits Pdx1 and MafA binding activities through different mechanisms: culture of isolated islets in the presence of palmitate prevents the nuclear translocation of Pdx1 that normally occurs upon glucose stimulation, while it blocks glucose induction of *MafA* mRNA expression[280]. The combined over-expression of Pdx1 and MafA prevents palmitate inhibition of *Insulin* gene expression in isolated rat islets, highlighting the essential role for these two transcription factors in the mechanisms of lipo-toxicity[280]. Basal K_{ATP} channel activity was 40% lower in islets cultured at 4.5mM and palmitate, and 60% lower in islets cultured at 15mM glucose plus either palmitate or oleate[281]. Insulin content decreased by 75% in islets exposed to FFAs in the presence of high (15mM) but not low (4.5mM) glucose concentrations, but the number of secretory granules was unchanged[281].

To induce β-cell dysfunction, I chose to treat cells with high glucose and palmitate. NIT-1 cells were treated in triplicate with 20 or 25mM glucose and 0.3 or 0.4mM palmitate, and GSIS assays were performed. The cells were starved for 2 hours at 37°C before incubation with 1xKRH for 1 hour (basal secetion) before incubation with 1XKRH with 20mM glucose for 1 hour (stimulated secretion). Released insulin in the supernatent was quantified with ELISA in technical triplicates. As Figure 40 shows, compared to cells grown under normal conditions, there is the most reduction in GSIS (60%) under conditions of 20mM glucose and 0.4mM palmitate. The increase in the GSIS response when the cells are treated with 20mM glucose compared to 15mM glucose has been observed before, and is probably due to the fact basal levels of insulin released is higher in cells treated with higher concentrations of glucose[282].

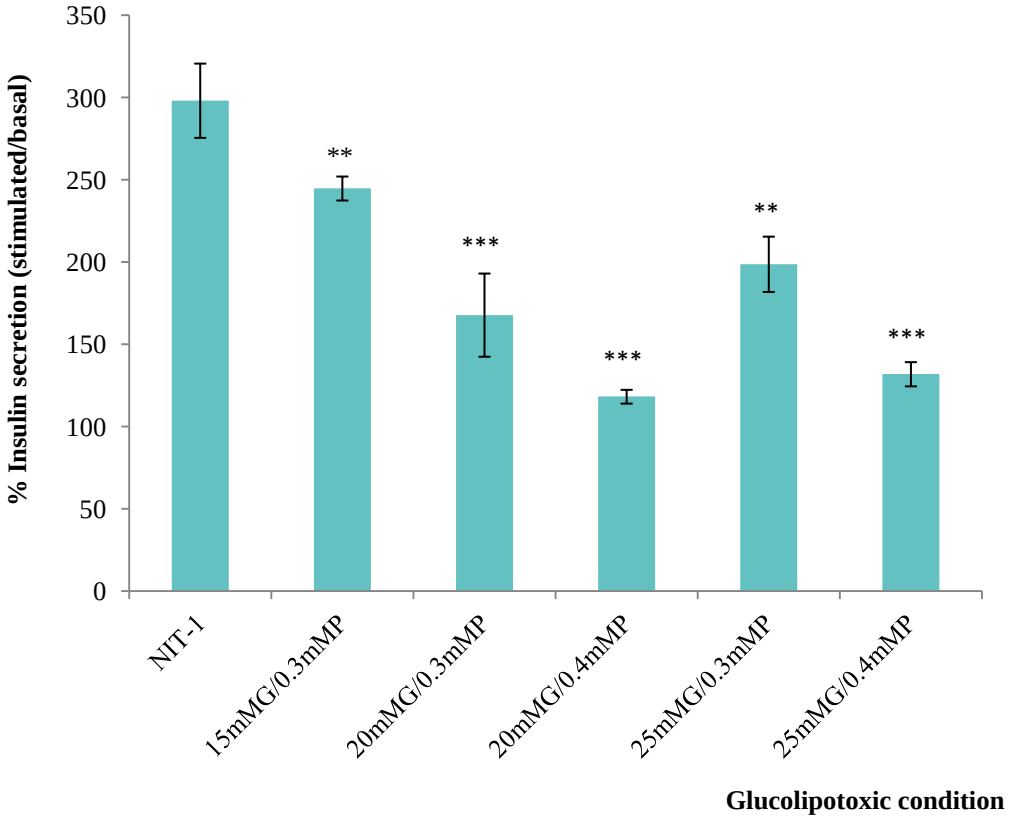


Figure 40. GSIS of NIT-1 cells treated with different concentrations of glucose and palmitate. NIT-1 cells were treated in triplicate with 20 or 25mM glucose (G) and 0.3 or 0.4mM palmitate (P), and GSIS assays were performed. The cells were starved for 2 hours at 37°C before incubation with 1xKRH for 1 hour (basal secetion) before incubation with 1XKRH with 20mM glucose for 1 hour (stimulated secretion). Released insulin in the supernatent was quantified with ELISA in technical triplicates and the values are means ± SE (n=3). ***$p \leq 0.01$, **$p \leq 0.02$. NIT-1 cells treated with 20mM glucose and 0.4mM palmitate showed the most decrease (60%) in the GSIS response.

5.2.2. Global Landscape of Activating and Repressive Regions of Chromatin in Normal and Dysfunctional Beta-cells

To look at H3K4me3 and H3K27me3 marks in normal and dysfunctional β-cells, I performed ChIP in normal NIT-1 β-cells and dysfunctional β-cells (with two conditions: cells treated with 15mM glucose and 0.3mM palmitate, and 20mM glucose and 0.4mM palmitate) with the H3K4me3 and H3K27me3 antibodies. When I performed the experiments in early 2010, there was no prior knowledge of where active and repressive regions of chromatin in β-cells were located. I therefore hypothesized that the *Ins1* gene would be active (i.e. H3K4me3-enriched), and that *MyoD* and *Utf1* (identified to not be bound by Pdx1, NeuroD1 nor MafA, Chapter 3) would be more repressed (i.e. H3K27me3-enriched). The *RIK2600* gene is bound by Pdx1 and NeuroD1 in normal NIT-1 cells, so I included that as a potential positive control for the H3K4me3 ChIP. Fold enrichment represents the abundance of enriched DNA fragments over a negative control region not bound by the antibody of interest, and I also reasoned that the regions highly enriched for H3K4me3 would probably not be marked by H3K27me3 and vice versa, so the regions highly enriched for H3K4me3 could serve as the negative control region for the H3K27me3 ChIP, and vice versa.

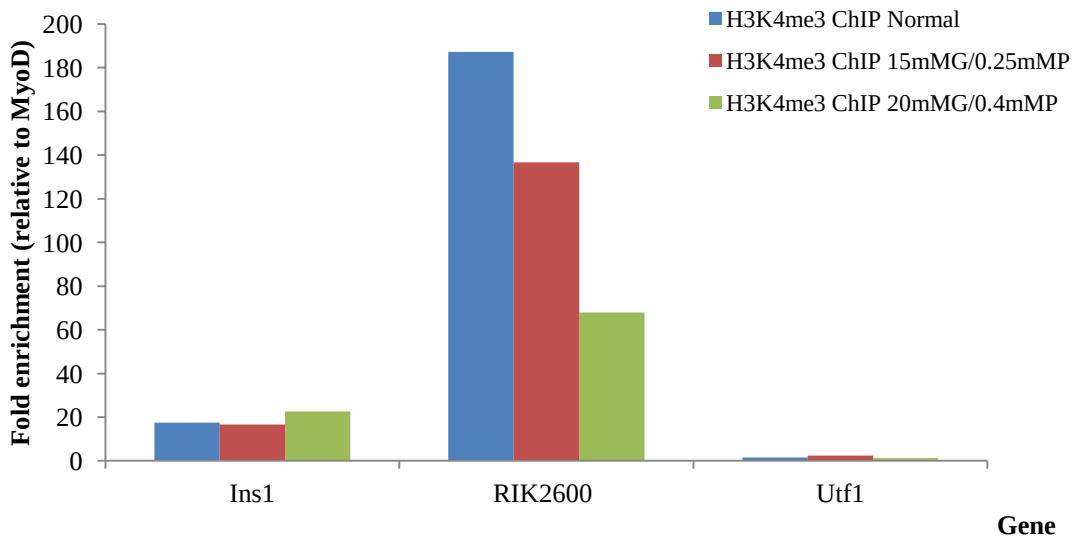
H3K4me3 was enriched at the *Ins1* gene in both normal and dysfunctional β-cells relative to *MyoD*, and the *RIK2600* gene showed a decrease in the level of H3K4me3 enrichment as the concentration of glucose and palmitate increased (Figure 41A). The other negative control region for the H3K4me3 ChIP was *Utf1*, and it showed no enrichment for H3K4me3 relative to *MyoD*. The converse was seen for the H3K27me3 ChIP, with low levels of enrichment for *Ins1* relative to *RIK2600* (the other negative control region for the H3K27me3 ChIP), and high enrichments for *MyoD* and *Utf1* (Figure 41B).

I extended the search to a genome-wide level, and performed ChIP-Seq with H3K4me3 and H3K27me3 on normal NIT-1 β-cells and dysfunctional β-cells. I selected dysfunctional β-cells treated with 20mM glucose and 0.4mM palmitate for subsequent genome-wide

experiments because the enrichments for H3K4me3 and H3K27me3 were higher in those conditions than in 15mM glucose and 0.3mM palmitate. There was also more of a decrease in the GSIS response in cells treated with 20mM glucose and 0.4mM palmitate (Figure 40). The ChIPs were performed in biological duplicates (i.e. chromatin was harvested on cells treated on two separate occasions with high glucose and fatty acids and ChIPs performed on separate days) and validated the quality of the ChIPs in the same fashion as in Figure 41 before preparing the sample for sequencing. Input DNA was sequenced as a control for background noise in both normal and dysfunctional β-cells, and subsequent reads mapped against that.

I also performed FAIRE on normal and dysfunctional β-cells to identify regions of open and closed chromatin. DNA segments that actively regulate transcription *in vivo* are typically characterized by eviction of nucleosomes from chromatin and are experimentally identified by their hypersensitivity to nucleases. In FAIRE, chromatin is cross-linked with formaldehyde *in vivo*, sheared by sonication, and phenol-chloroform extracted. Cross-linking between histones and DNA (or between one histone and another) is likely to dominate the chromatin crosslinking profile[283,284]. Therefore, covalently linked protein–DNA complexes are sequestered to the organic phase, leaving only protein-free DNA fragments in the aqueous phase, which is then recovered and prepared for sequencing. For the background reference, the same procedure is performed on a portion of the cells that had not been fixed with formaldehyde (identical to a traditional phenol-chloroform extraction), and sequenced.

(A)



(B)

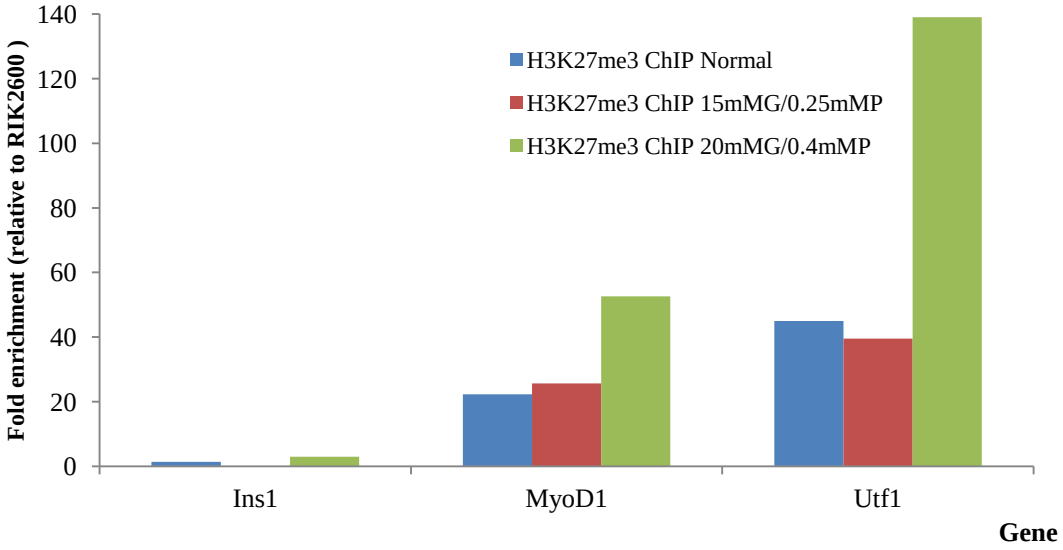


Figure 41. ChIP of H3K4me3 and H3K27me3 in normal NIT-1 β-cells and β-cells treated with varying concentrations of glucose and palmitate. Chromatin from NIT-1 cells in different treatment conditions: normal media (blue bars), 15mM glucose and 0.25mM palmitate (red bars) and 20mM glucose and 0.4mM palmitate (red bars) was sheared and ChIP was performed with antibodies against (A) H3K4me3 and (B) H3K27me3. Fold enrichment represents the abundance of enriched DNA fragments over a control region. Each ChIP experiment was performed with three technical replicates for each RT-PCR, and the average of the normalized ratio to the target gene/control region was calculated and presented. This data is representative of ChIP experiments performed in three biological replicates.

In analysis of the H3K4me3 and H3K27me3 ChIP-Seq libraries, raw mapped reads were compiled for each biological duplicate and HOMER[189] was used for peak calling. Control samples (genome sequencing without antibodies for H3K4me3 and H3K27me3) were used as the background input, with two-fold enrichment relative to the background and FDR 0.001 as the thresholds. For FAIRE analysis, F-Seq[190] was used for peak identification using the raw mapped reads. The sequencing output of DNA without formaldehyde fixation was used for the control, and the identified peaks from the control sequencing were subtracted from those in FAIRE sequencing. The standard deviation of 6.0 was used for the threshold. The closest TSS from each peak was identified for gene annotation.

There was a notably larger number of FAIRE peaks for dysfunctional β-cells than normal cells, 61962 in contrast to 46668 (Table 10). There was also an increase in the number of H3K4me3 ChIP-Seq peaks in dysfunctional β-cells when compared to normal β-cells, 15780 peaks vs. 14204. There was a decrease in H3K27me3 peaks in dysfunctional β-cells, 10739 compared to 12686 peaks in normal β-cells. At the promoter regions, there were also more FAIRE peaks (5720) and H3K4me3 (3117) peaks in dysfunctional β-cells than normal cells, which had 341 FAIRE peaks and 2892 H3K4me3 peaks. There was a decrease in the number of H3K27me3 peaks located at the promoter in dysfunctional β-cells (555) compared to normal β-cells (664). Taken together, it suggests that there is an increase in the global level of open chromatin at the promoters of dysfunctional β-cells when compared to normal β-cells.

The average profile of H3K4me3 was plotted across the TSS and there was a binomial distribution of peaks at the TSS (Figure 42). The peak downstream of the TSS was higher, compared to the peak before the TSS, this is consistent with the profile of H3K4me3 normally seen at promoters[136], and probably demarcates transcription factor binding. There was a dip in peaks at the end of the TSS, also consistent with previous reports[136]. The average profile of H3K27me3 was plotted across the TSS and there was a single peak at the TSS (Figure 43).

	Normal			Dysfunctional		
	FAIRE	**H3K4me3**	**H3K27me3**	**FAIRE**	**H3K4me3**	**H3K27me3**
# total peaks	46668	14204	12686	61962	15780	10739
# promoter peaks	341	2892	664	5270	3117	555
# TSS peaks	347	259	115	458	268	80
# 5'UTR peaks	6	566	76	92	612	47
# intergenic peaks	14306	1205	4972	7657	1820	4150

Table 10. Summary of mapped reads and peak calling of FAIRE, H3K4me3 and H3K27me3 ChIP-Seq libraries in normal and dysfunctional β-cells.

Raw mapped reads (in BED format) were compiled for each biological duplicate and were used for peak calling by HOMER[189]. Control samples (genome sequencing without antibodies for H3K4me3 and H3K27me3) were used as the background input, with two-fold enrichment relative to the background and an FDR of 0.001 were used as the thresholds. For the FAIRE experiments, F-Seq[190] was used for peak identification based on the raw mapped reads. The sequencing output of DNA without formaldehyde fixation was used for control. The identified peaks from the control sequencing were subtracted from those in FAIRE sequencing. (Dr Choi performed this analysis.)

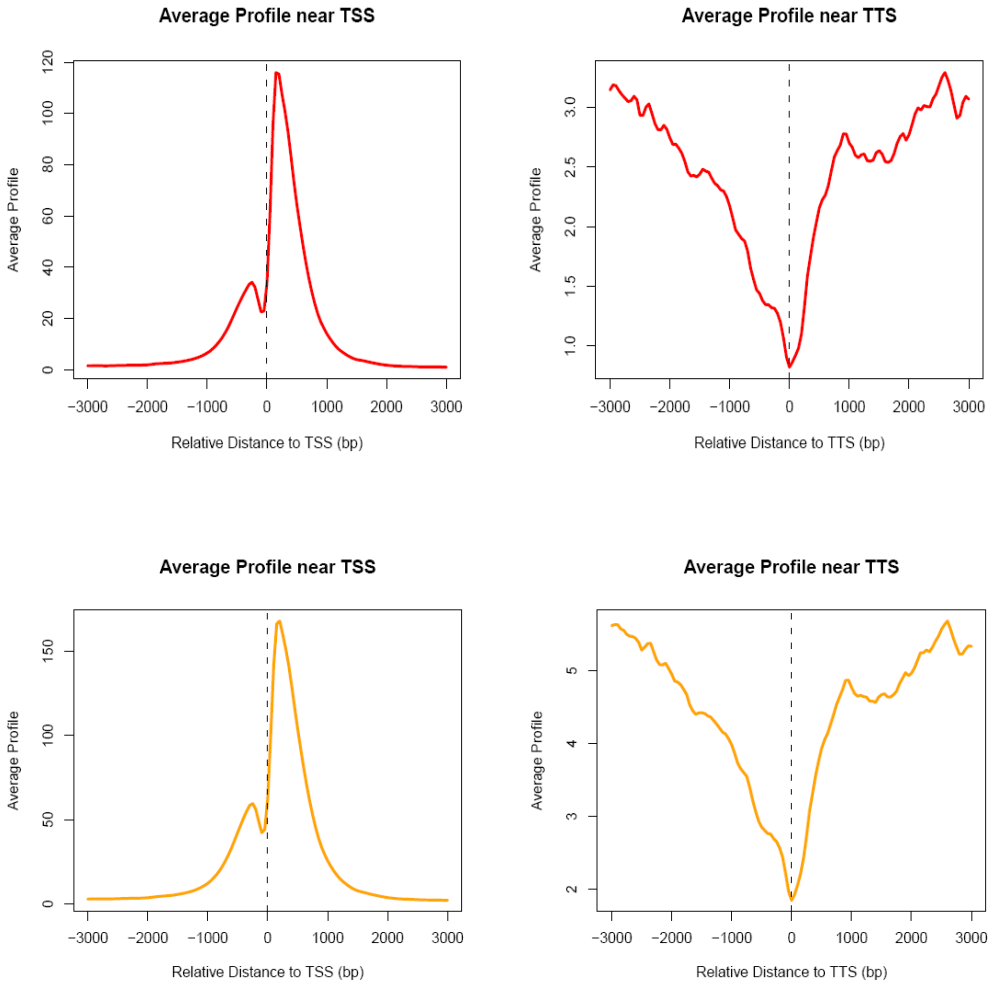


Figure 42. H3K4me3 methylation at TSS. The average profile of H3K4me3 was plotted across the TSS and there was a binomial distribution of peaks at the TSS. Y axis: modification level or DNA-binding level. (Dr Choi performed this analysis.)

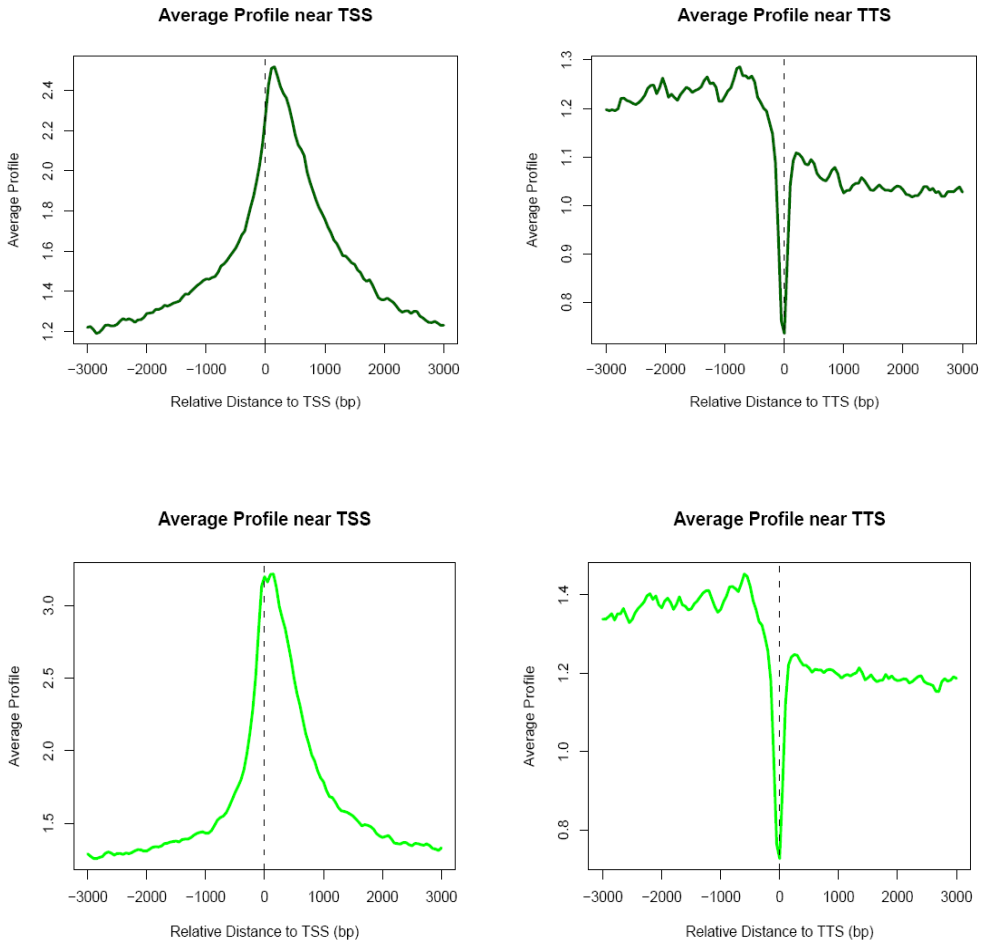


Figure 43. H3K27me3 methylation at TSS. The average profile of H3K27me3 was plotted across the TSS and there was a distribution of peaks at the TSS. (Dr Choi performed this analysis.)

It was interesting that there was more open chromatin in dysfunctional β-cells especially in the FAIRE libraries. I hypothesized there were differences in the *cis*-regulatory elements found at the open chromatin regions, so motif finding was performed with these libraries. The motif lengths of 8, 10, and 12bp were searched against the repeat-masked mm8 *Mus musculus* genome and position-specific weight matrices were calculated based on the HOMER database. Motif analysis of open chromatin in normal cells show top enrichments for the transcription factors FoxA1 (p value = 1.15E-42), FoxA2 (p value = 8.27E-42), POU class 5 homeobox 1 (Oct)-4 (p value = 7.06E-32), POU class 3 homeobox (Oct)-2 (p value = 1.59E-30), and Pdx1 (p value = 7.78E-29) (Table 11A). The transcription factor Upstream stimulating factor (Usf)-1 was also in the top 20 enriched motifs, with a p value of 2.24E-06. In contrast, open chromatin of dysfunctional β-cells show top enrichments for general regulators such as Sp transcription factor (Sp)-1 (p value = 1.09E-109), CTCF (p value = 1.42E-54) and Nfy(CCAT)-binding factor (p value = 4.73E-53) (Table 11B). A full list of motifs are listed in Supplementary Table 7.

All these indicate that in dysfunctional cells, open chromatin is biased to binding sites for general transcription factors, while in normal cells, tissue-specific TF binding sites tend to have a more open chromatin configuration.

(A)

Motif Name	p-value
FOXA1(Forkhead)/LNCAP-FOXA1-ChIP-Seq/Homer	1.15E-42
Foxa2(Forkhead)/Liver-Foxa2-ChIP-Seq/Homer	8.27E-42
Oct4(POU/Homeobox)/mES-Oct4-ChIP-Seq/Homer	7.06E-32
Oct2(POU/Homeobox)/Bcell-Oct2-ChIP-Seq/Homer	1.59E-30
Pdx1(Homeobox)/Islet-Pdx1-ChIP-Seq/Homer	7.78E-29

(B)

Motif Name	p-value
Sp1(Zf)/Promoter/Homer	1.09E-199
Klf4(Zf)/mES-Klf4-ChIP-Seq/Homer	1.90E-98
CTCF(Zf)/CD4+-CTCF-ChIP-Seq/Homer	1.42E-54
NFY(CCAAT)/Promoter/Homer	4.73E-53
NRF1/Promoter/Homer	2.06E-36

Table 11. FAIRE peaks in (A) normal β-cells show tissue-specific transcription factor motifs, while (B) motifs in dysfunctional β-cells show enrichment for general transcription factors.

The motif lengths of 8, 10, and 12bp were searched against the repeat-masked mm8 *Mus musculus* genome and position-specific weight matrices were calculated based on the HOMER's database. (Dr Choi performed this analysis.)

5.2.3. Genes Epigenetically Altered in Dysfunctional Beta-Cells

I next wanted to know if the genes associated with the respective epigenetic marks in normal and dysfunctional β-cells had similarities in their biological networks and/or molecular roles. To achieve this, I uploaded the RefSeq identity list of nearest associated genes of each chromatin mark in each cellular condition (H3K4me3, H3K27me3, FAIRE in normal and dysfunctional β-cells) into IPA[187] and identified the top biological networks and cellular processes and functions associated with them. I then utilized the comparison function in IPA to compare the matching datasets in the two conditions (genes marked by FAIRE in normal β-cells vs. genes marked by FAIRE in dysfunctional β-cells, and the same for H3K4me3 and H3K27me3).

There was more of an over-representation of genes identified by FAIRE-Seq that are involved in nervous system development in normal β-cells compared to dysfunctional β-cells, and an over-representation of genes involved in connective tissue development and embryonic development in dysfunctional β-cells (Figure 44). In the FAIRE-Seq dataset there are also genes more enriched for infectious mechanisms and infectious diseases in dysfunctional β-cells compared to normal β-cells. It is of interest that there was a higher enrichment for genes involved in cellular growth and proliferation, cell death, cell cycle, molecular transport, DNA replication, recombination and repair, protein synthesis and gene expression in dysfunctional β-cells when compared to normal β-cells. There were also genes involved in protein trafficking enriched only in the dysfunctional β-cells, while genes involved in endocrine system development and function were enriched only in the normal β-cells. This shows that genes identified to be in regions of open chromatin by FAIRE-Seq in dysfunctional β-cells are more enriched for in general cellular regulation and development of other developmental lineages, including embryonic development, when compared to normal β-cells, in agreement with the motifs identified in Section 5.2.2. A full list of genes involved in those pathways is listed in Supplementary Table 8A.

There was an over-representation of genes involved in cellular development, cardiovascular development and function, embryonic development, nervous system and development, organismal development, skeletal and muscular system development and function, tissue development, and respiratory development and function, in the genes enriched for H3K4me3 in dysfunctional β-cells when compared to normal β-cells (Figure 45). Similar to the FAIRE-Seq dataset, there was also an over-representation of genes involved in cellular growth and proliferation, cellular movement, and lipid metabolism in dysfunctional β-cells. Disease-related genes found to be more enriched in the dysfunctional β-cells were those involved in endocrine system disorders, gastrointestinal diseases, genetic disorders, metabolic disease, neurological disease, cardiovascular disease, and psychological disorders. In line with the comparison of genes identified by FAIRE-Seq, these results show that genes identified to be in regions of active chromatin by H3K4me3-Seq in dysfunctional β-cells are more involved in general cell regulatory activities and the development of other lineages, including embryonic development, when compared to normal β-cells. A full list of genes involved in those pathways is listed in Supplementary Table 8B.

Genes implicated in genetic disease, skeletal and muscular disorders, neurological disorders, connective tissue disorders, inflammatory disease, cardiovascular and haematological disease were more enriched for H3K27me3 genes in normal β-cells than dysfunctional β-cells. Genes involved in carbohydrate metabolism were slightly more enriched in the H3K27me3 genes in dysfunctional β-cells when compared to normal β-cells (Figure 46). Similarly, genes implicated in nutritional disease were only identified to be significantly enriched for H3K27me3 genes in dysfunctional β-cells and not in normal β-cells. There are more genes enriched for H3K27me3 in normal β-cells involved in system development and function, namely the nervous system development and function, renal and urological, skeletal and muscular, cardiovascular, embryonic development. There was an over-representation of genes with H3K27me3 involved in cell morphology, cell-to-cell signalling and interaction and cell death, in normal β-cells when compared to dysfunctional β-cells. Interestingly, there are more

genes associated with H3K27me3 involved in organismal injury and abnormalities in dysfunctional β-cells than normal cells. This shows that genes identified to be in repressive regions of chromatin by H3K27me3-Seq in normal β-cells are enriched for the development and function of different lineages, various diseases, and cell regulation. A full list of genesinvolved in those pathways is listed in Supplementary Table 8C.

Taken together, the comparison of the genes associated with the respective epigenetic marks in normal and dysfunctional β-cells demonstrates that there is a significant increase in genes involved in cellular regulatory activities and the specification of other developmental lineages in dysfunctional β-cells.

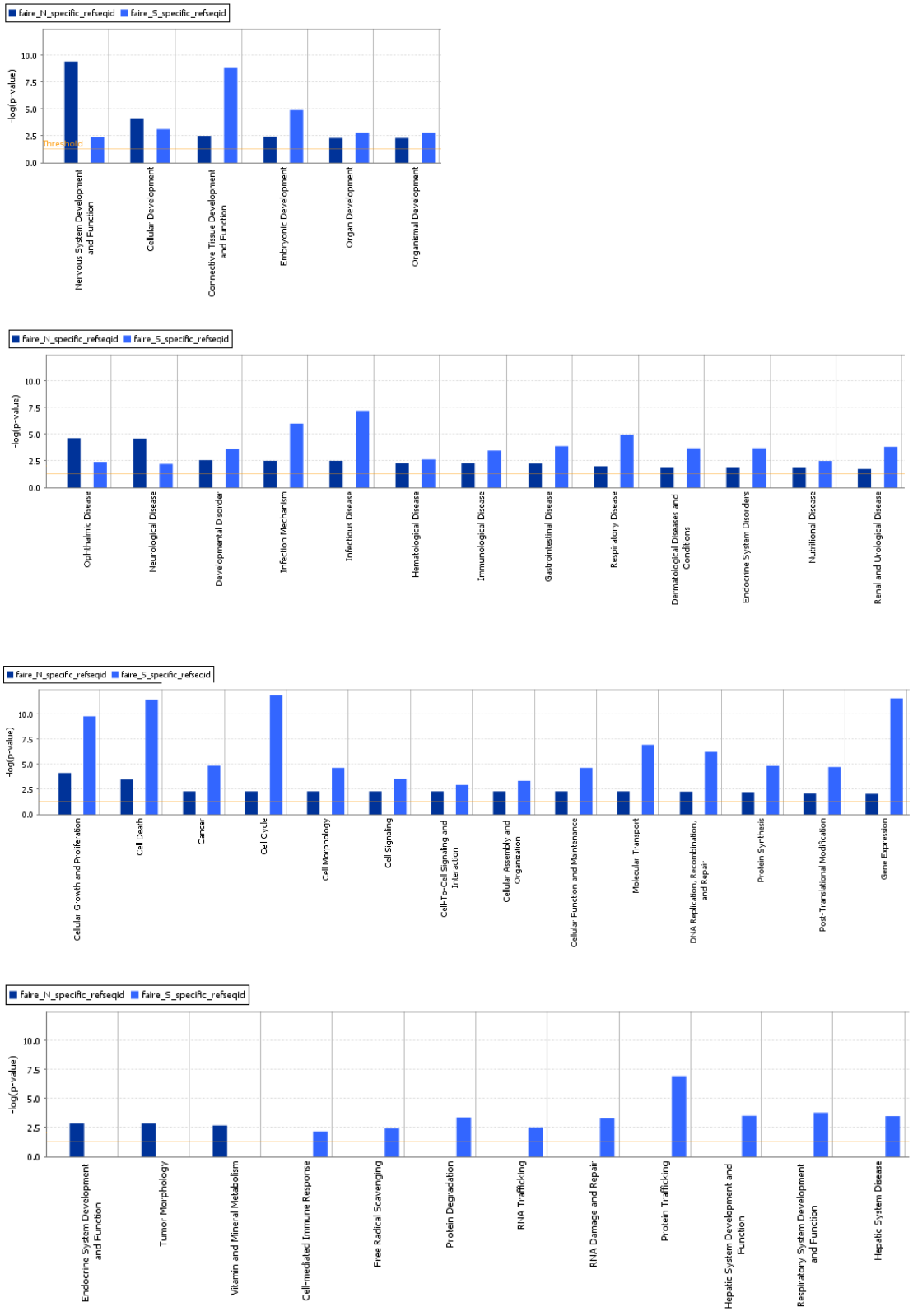


Figure 44. Comparison of FAIRE-peaks in normal and dysfunctional β-cells with IPA shows more enrichment of genes involved in general cellular regulation and development of other developmental lineages in dysfunctional β-cells. N= normal β-cells, S= dysfunctional β-cells.

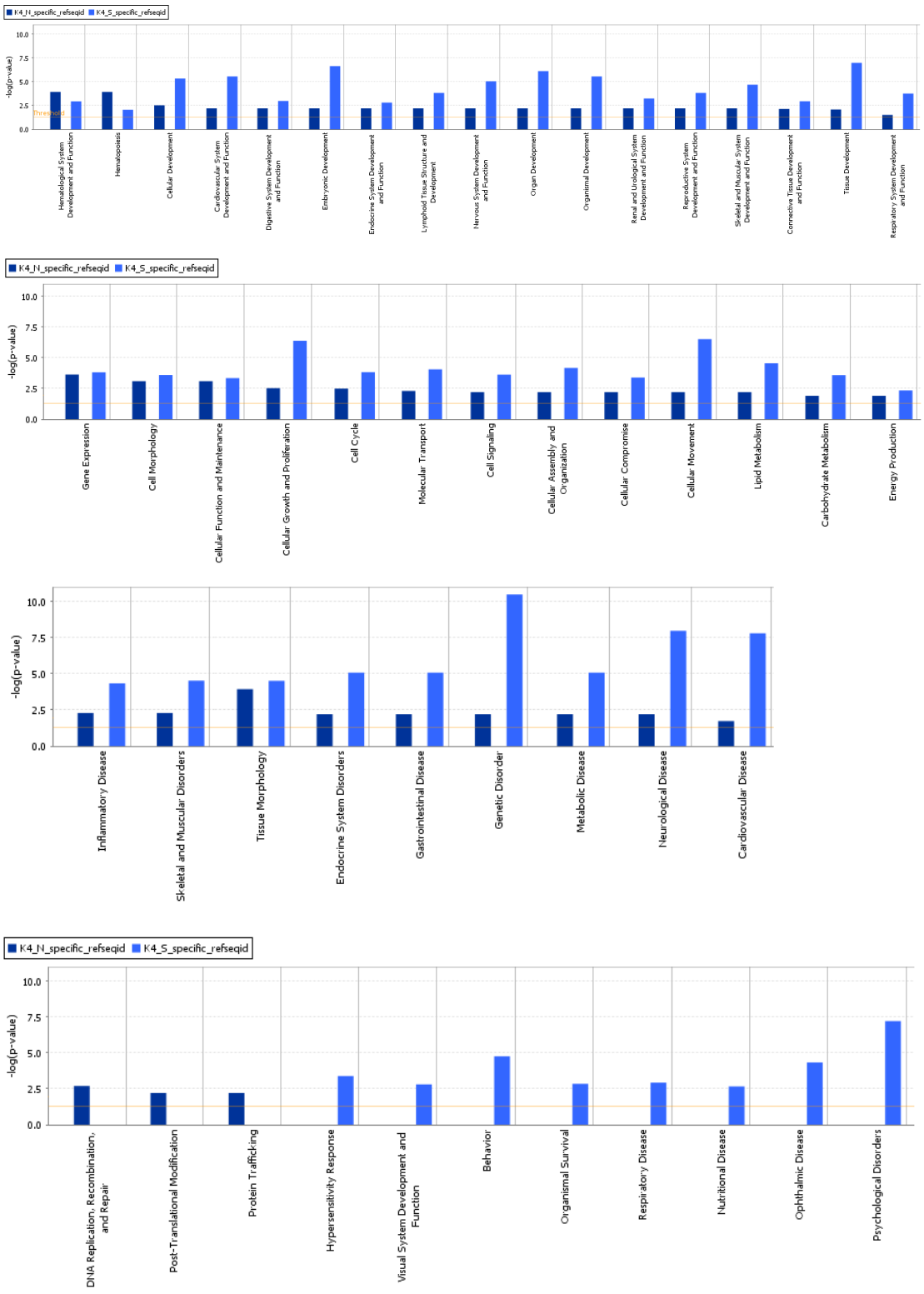
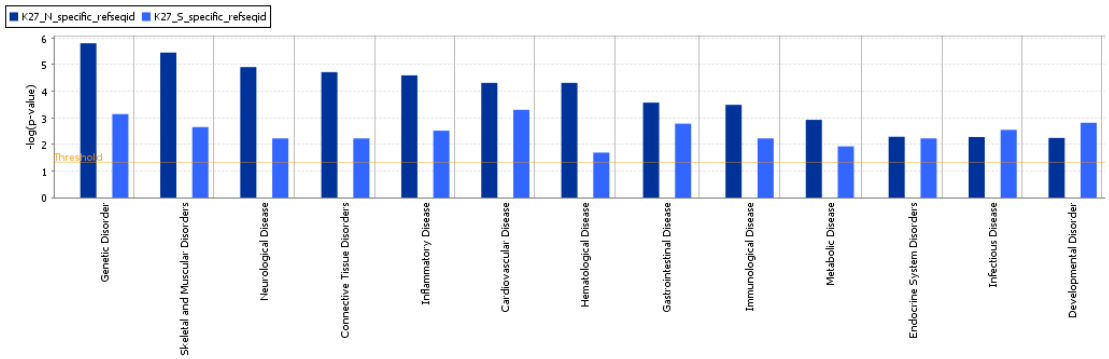
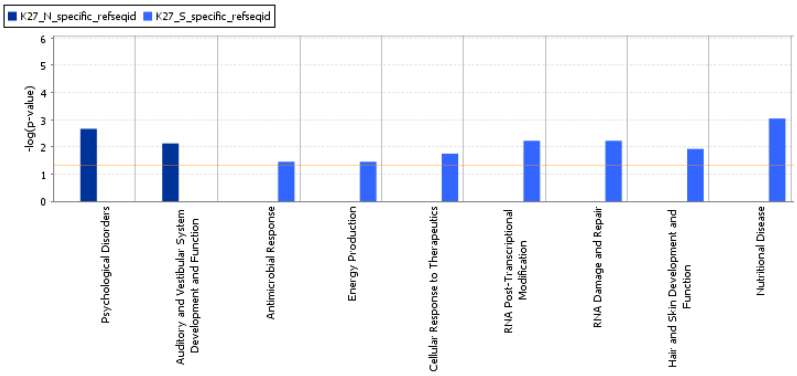
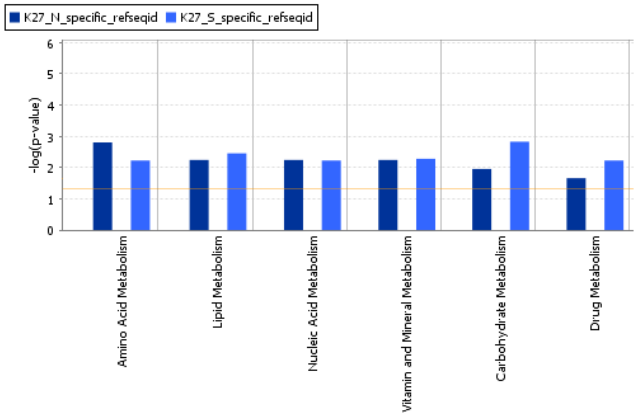
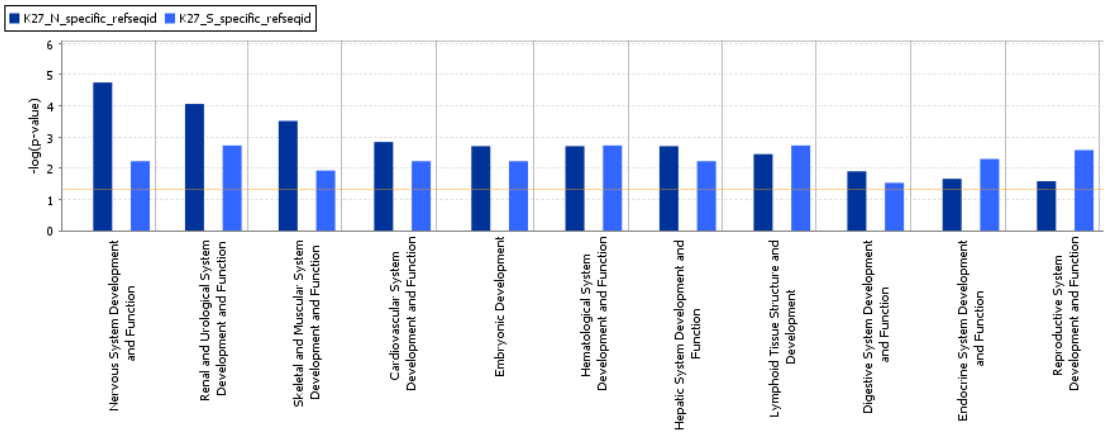


Figure 45. Comparison of H3K4me3-peaks in normal and dysfunctional β-cells with IPA shows more enrichment of genes involved in general cellular regulation and development of other developmental lineages in dysfunctional β-cells. N= normal β-cells, S= dysfunctional β-cells.


K27_N_specific_refseqid
K27_S_specific_refseqid
-log(p-value)
Threshold
Genetic Disorder
Skeletal and Muscular Disorders
Neurological Disease
Connective Tissue Disorders
Inflammatory Disease
Cardiovascular Disease
Hematological Disease
Gastrointestinal Disease
Immunological Disease
Metabolic Disease
Endocrine System Disorders
Infectious Disease
Developmental Disorder


K27_N_specific_refseqid
K27_S_specific_refseqid
-log(p-value)
Psychological Disorders
Auditory and Vestibular System Development and Function
Antimicrobial Response
Energy Production
Cellular Response to Therapeutics
RNA Post-Transcriptional Modification
RNA Damage and Repair
Hair and Skin Development and Function
Nutritional Disease


K27_N_specific_refseqid
K27_S_specific_refseqid
-log(p-value)
Amino Acid Metabolism
Lipid Metabolism
Nucleic Acid Metabolism
Vitamin and Mineral Metabolism
Carbohydrate Metabolism
Drug Metabolism


K27_N_specific_refseqid
K27_S_specific_refseqid
-log(p-value)
Nervous System Development and Function
Renal and Urological System Development and Function
Skeletal and Muscular System Development and Function
Cardiovascular System Development and Function
Embryonic Development
Hematological System Development and Function
Hepatic System Development and Function
Lymphoid Tissue Structure and Development
Digestive System Development and Function
Endocrine System Development and Function
Reproductive System Development and Function

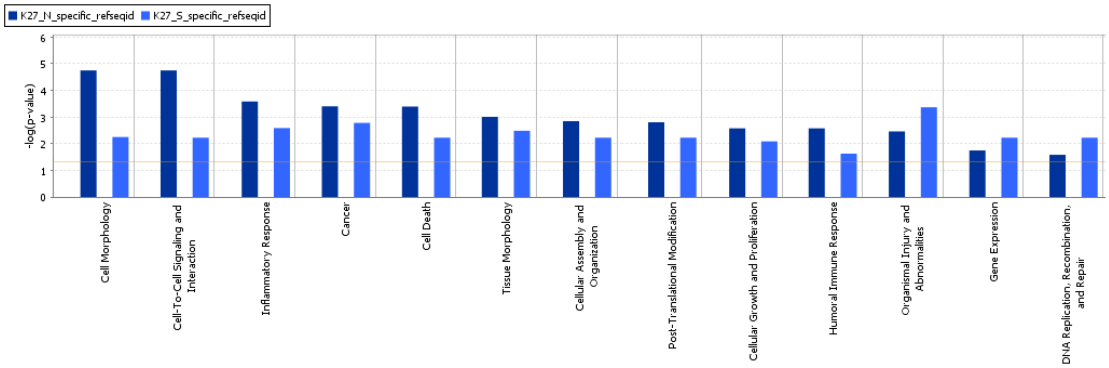
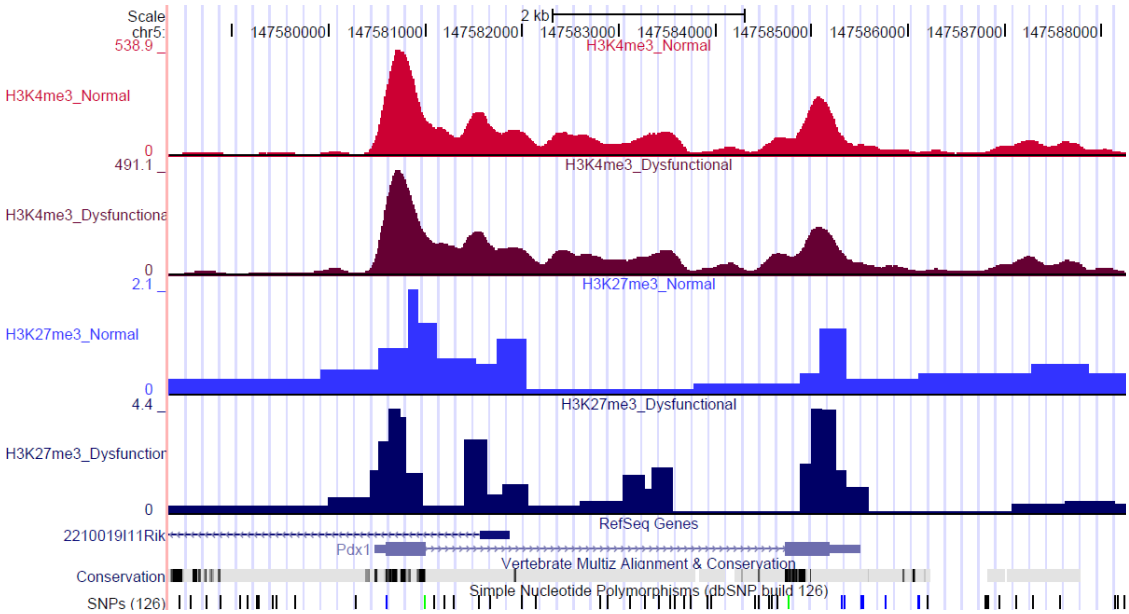


Figure 46. Comparison of H3K27me3-peaks in normal and dysfunctional β-cells with IPA shows more enrichment of genes involved in the development and function of different lineages, various diseases, and cell regulation in normal β-cells. N= normal β-cells, S= dysfunctional β-cells.

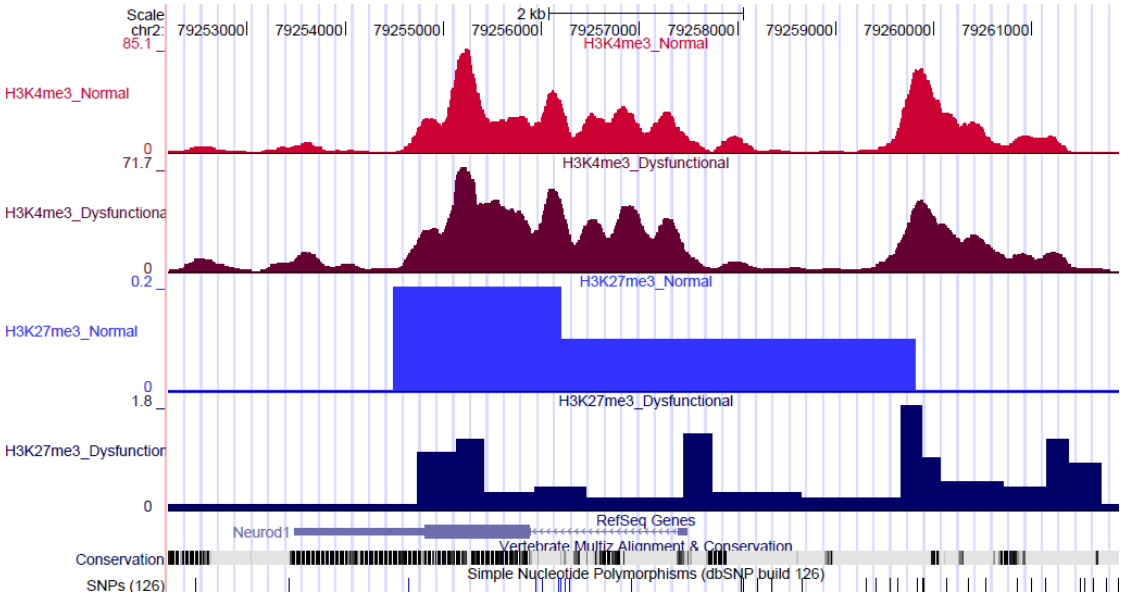
5.2.4. Dysfunctional Beta-cells Are Repressively Marked at Important β-cell Genes

I next wanted to see if specific genes important for normal β-cell function had differential chromatin profiles in dysfunctional β-cells, so ChIP-Seq peaks were visualized in the UCSC genome browser with the NCBI36/mm8 *Mus musculus* reference assembly. As shown in Figure 47, (A) *Pdx1*, (B) *NeuroD1*, (C) *MafA* are marked by less H3K4me3 in dysfunctional β-cells than in normal cells, as evidenced by the decrease in number of tags in the y axis. This was coupled with an increase in the number of H3K27me3 tags at the same locations. Surprisingly, *Ins1* (Figure 47D) had a slight decrease in the number of H3K4me3 tags at the promoter. I next looked at some genes important for the response of β-cells to a glucose stimulus (highlighted in Chapter 3, Table 9). There was less H3K4me3 at *Glut2*, *Snap25*, *Vamp2*, *Cacna1a*, *Chgb*, *Kirrel6.2*, *Sytl4*, *Stxbp1*, and *Syt7*, coupled with an increase in H3K27me3 (Figure 47E-M). I also looked at genes involved in other β-cell functions, and *Ucp2* (Figure 47N), and *Usf1* (Figure 47O) had the same pattern of chromatin remodelling.

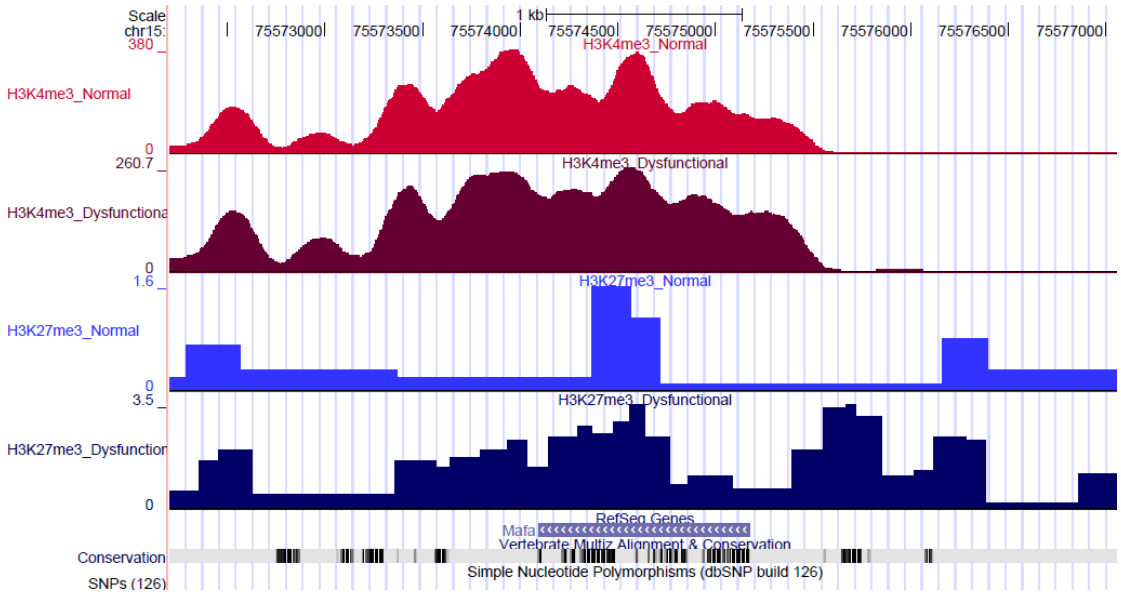
(A) Pdx1



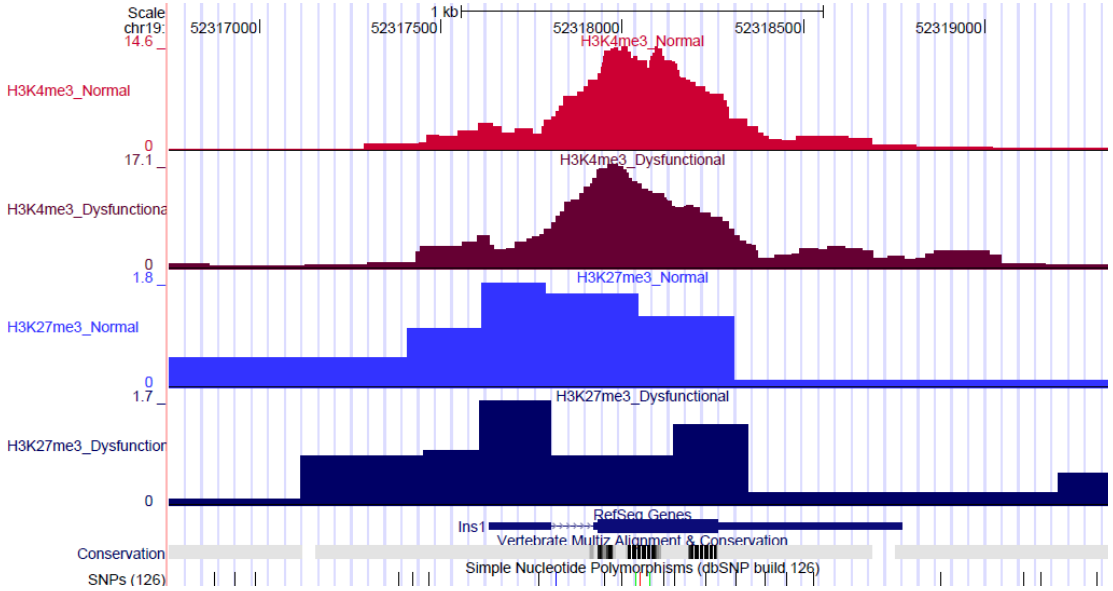
(B) NeuroD1



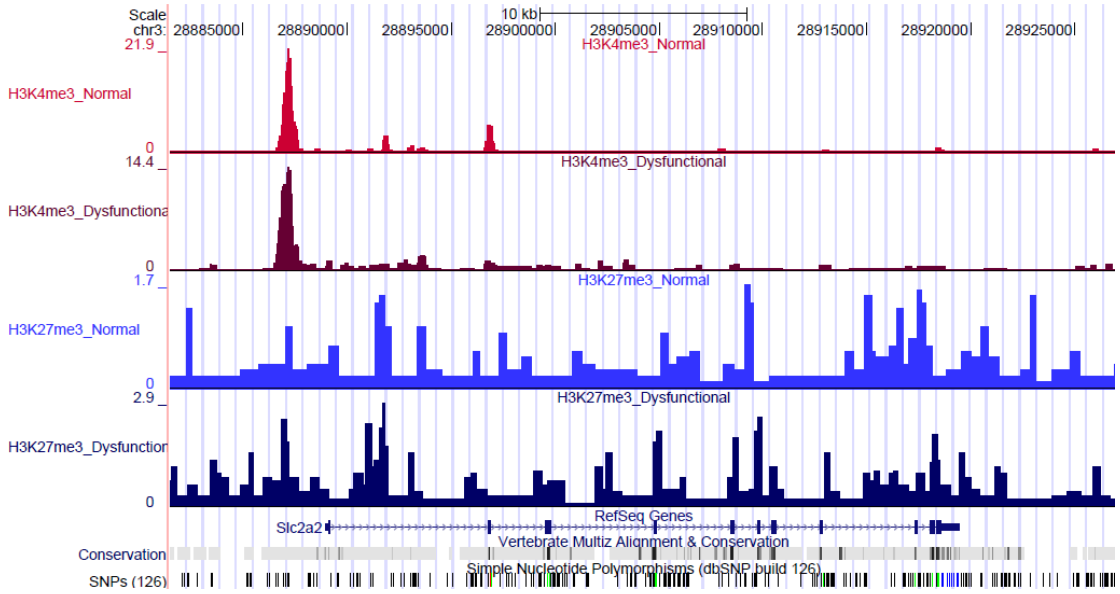
(C) MafA


Scale
chr15:
1 kb
75573000
75573500
75574000
75574500
75575000
75575500
75576000
75576500
75577000
380
H3K4me3_Normal
0
260.7
H3K4me3_Dysfunctional
0
1.6
H3K27me3_Normal
0
3.5
H3K27me3_Dysfunctional
0
RefSeq Genes
Mafa
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

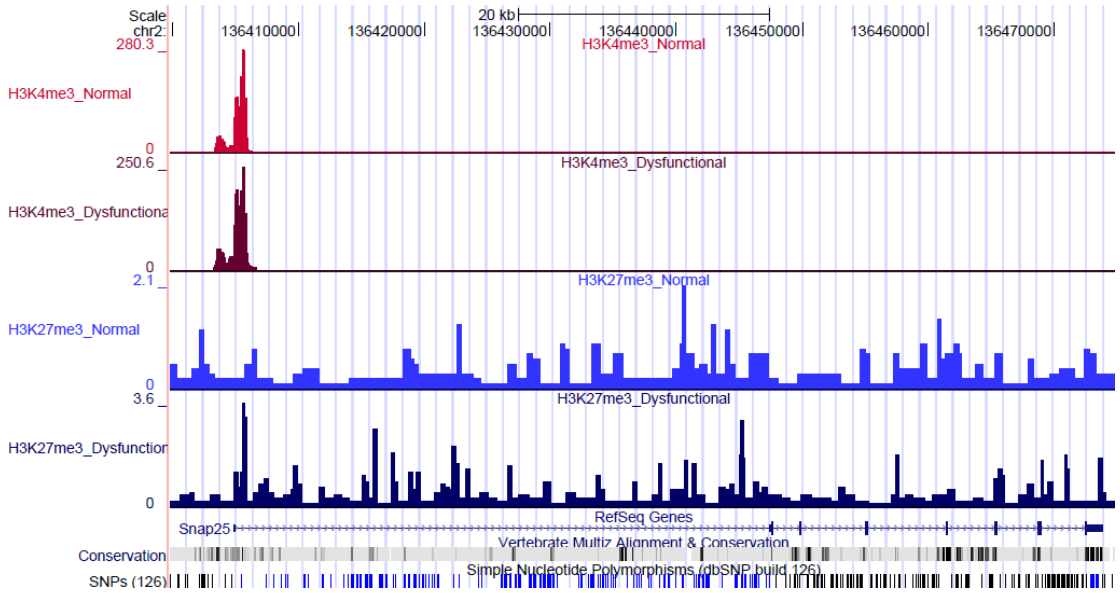
(D) Ins1


Scale
chr19:
1 kb
52317000
52317500
52318000
52318500
52319000
14.6
H3K4me3_Normal
0
17.1
H3K4me3_Dysfunctional
0
1.8
H3K27me3_Normal
0
1.7
H3K27me3_Dysfunctional
0
RefSeq Genes
Ins1
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

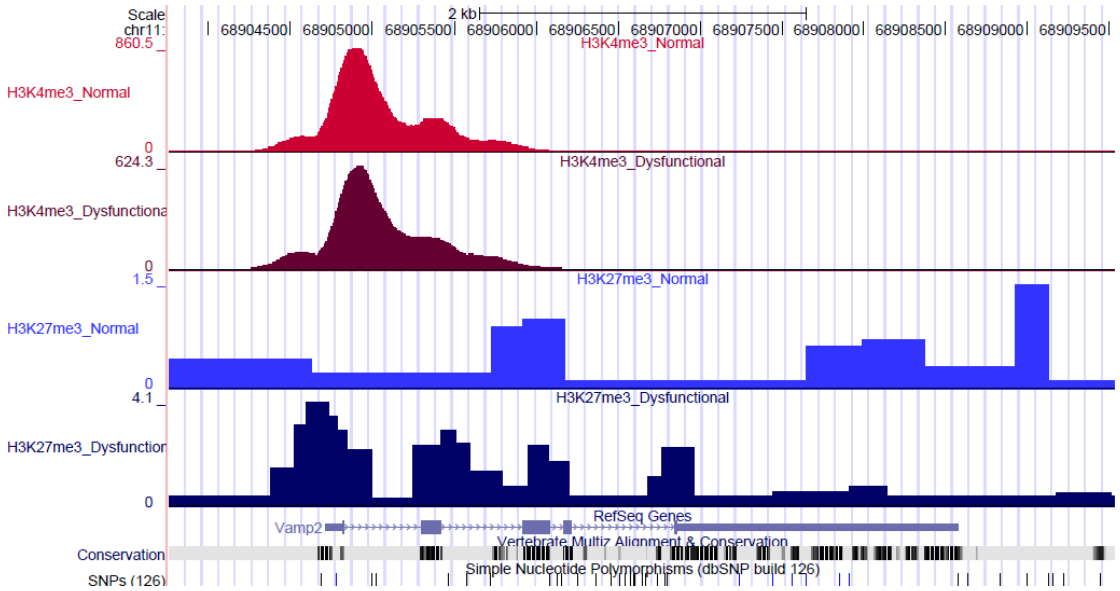
(E) Slc2a2



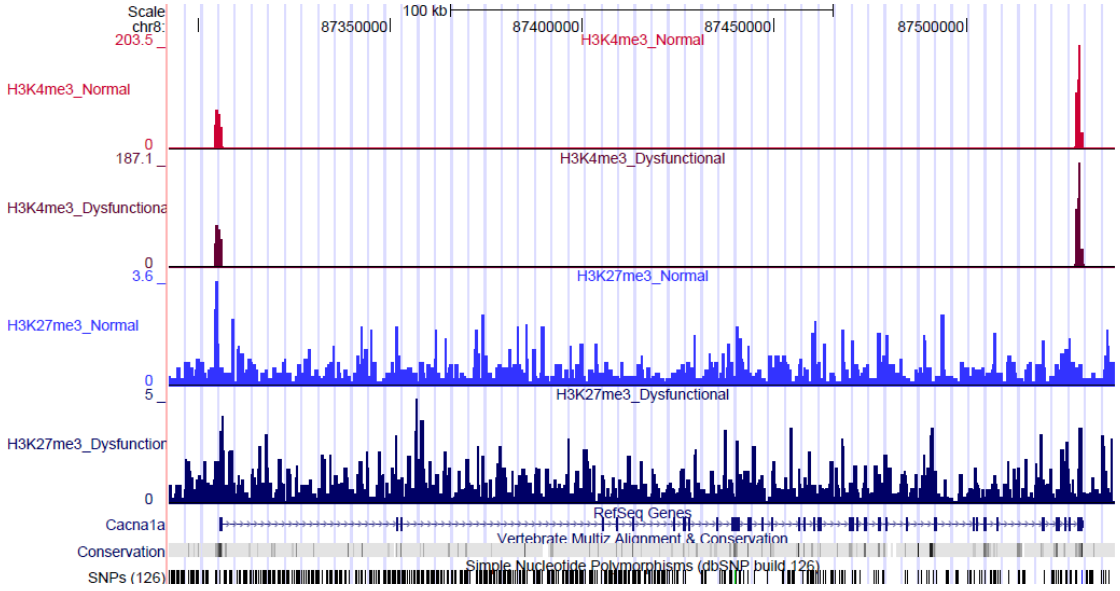
(F) Snap25



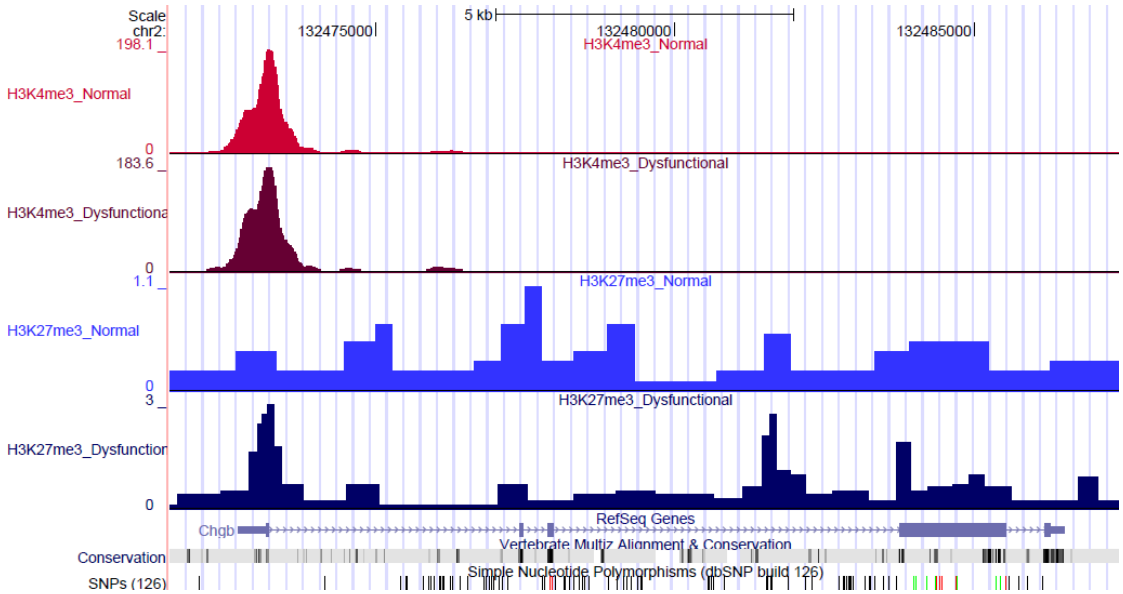
(G) Vamp2



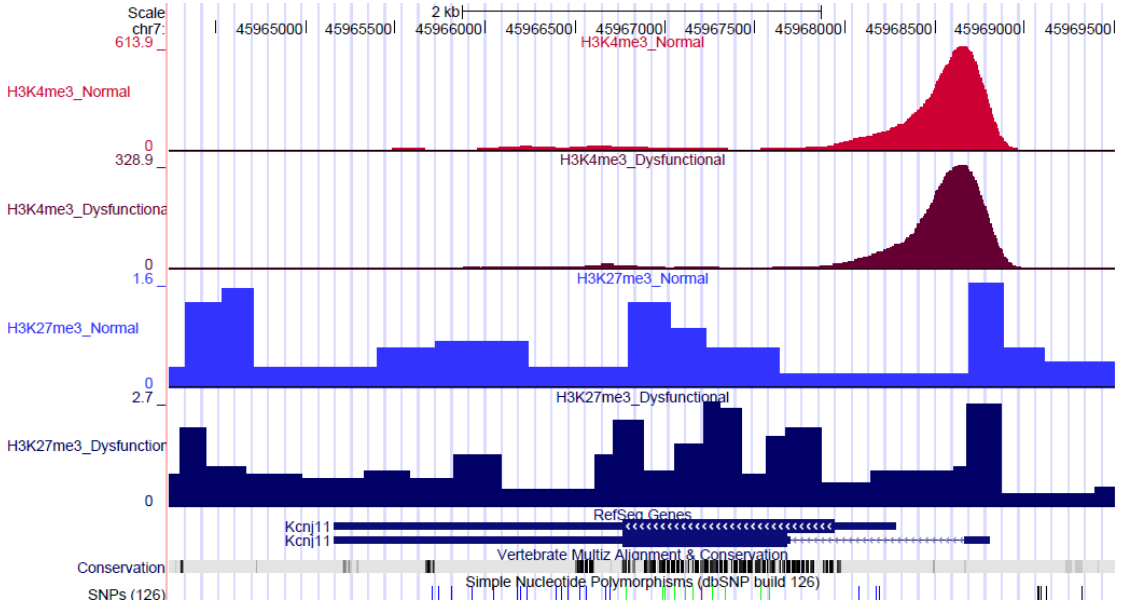
(H) Cacna1a



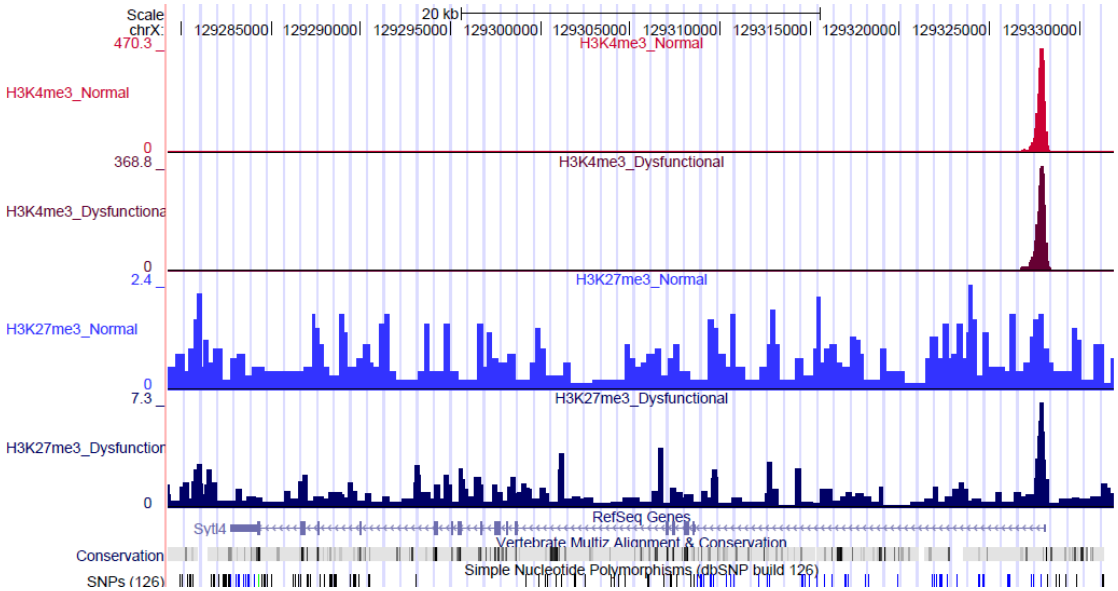
(I) Chgb



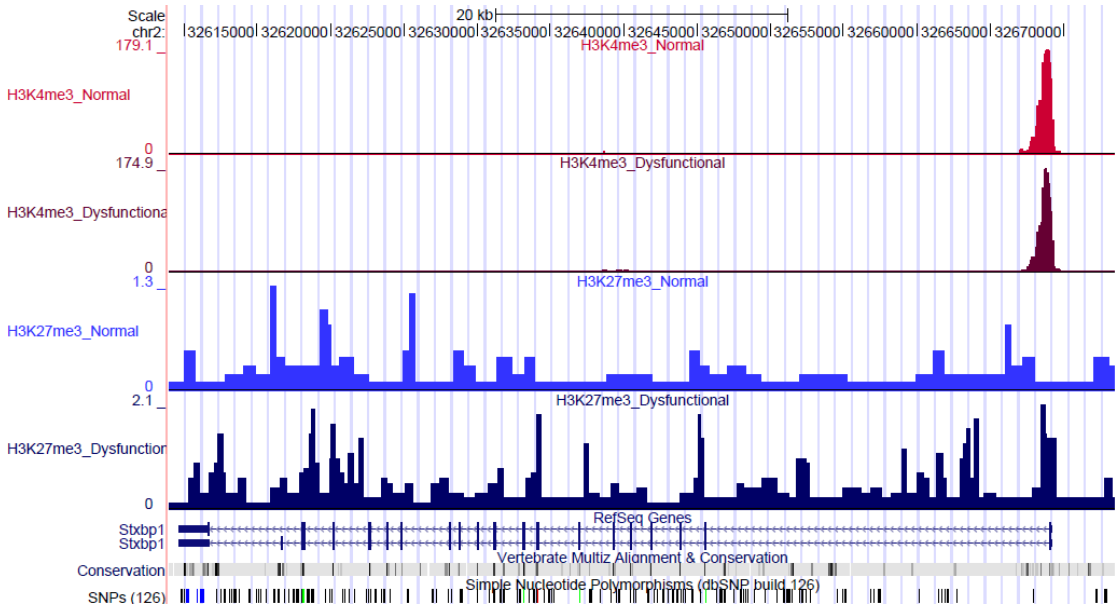
(J) Kirrel6.2



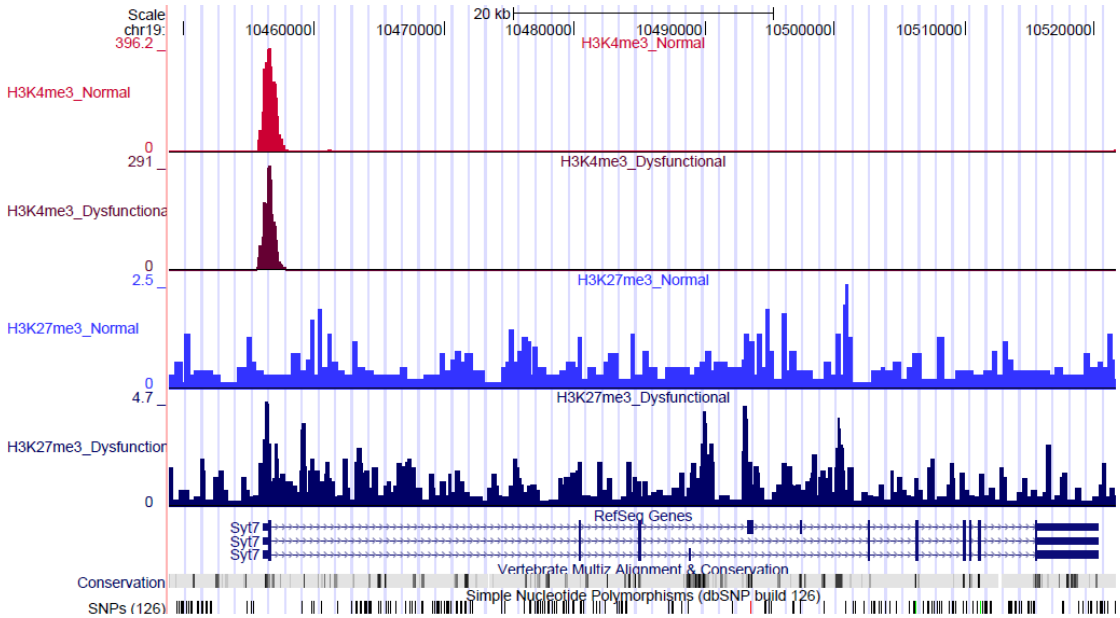
(K) Sytl4



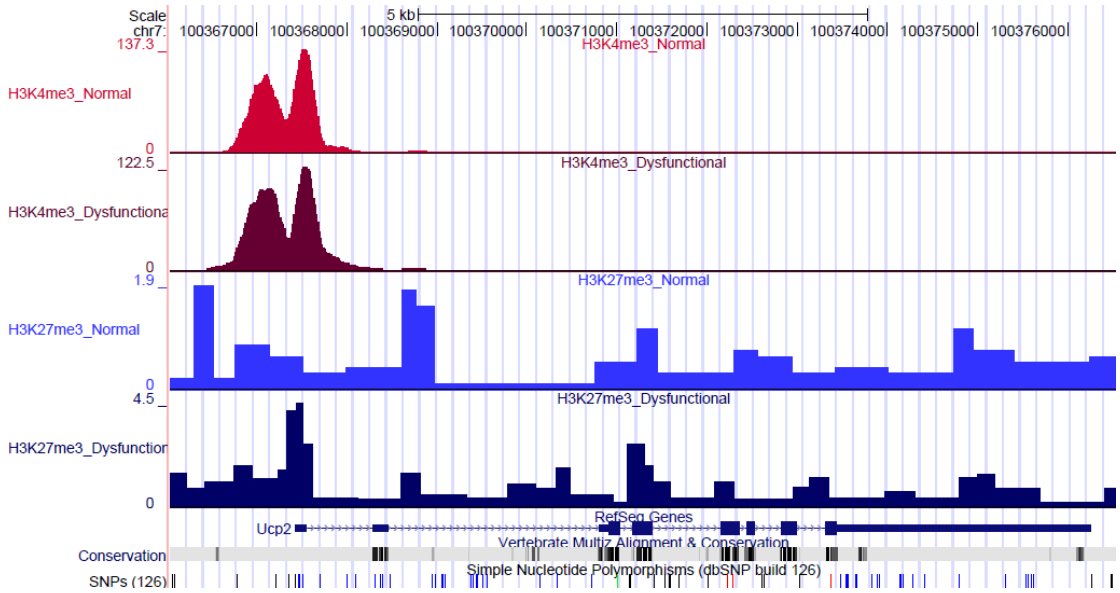
(L) Stxbp1



(M) Syt7


Scale
chr19:
20 kb
10460000
10470000
10480000
10490000
10500000
10510000
10520000
396.2
H3K4me3_Normal
H3K4me3_Normal
0
291
H3K4me3_Dysfunctional
H3K4me3_Dysfunctiona
0
2.5
H3K27me3_Normal
H3K27me3_Normal
0
4.7
H3K27me3_Dysfunctional
H3K27me3_Dysfunctior
0
RefSeq Genes
Syt7
Syt7
Syt7
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

(N) Ucp2


Scale
chr7:
5 kb
100367000
100368000
100369000
100370000
100371000
100372000
100373000
100374000
100375000
100376000
137.3
H3K4me3_Normal
H3K4me3_Normal
0
122.5
H3K4me3_Dysfunctional
H3K4me3_Dysfunctiona
0
1.9
H3K27me3_Normal
H3K27me3_Normal
0
4.5
H3K27me3_Dysfunctional
H3K27me3_Dysfunctior
0
RefSeq Genes
Ucp2
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

(O) Usf1

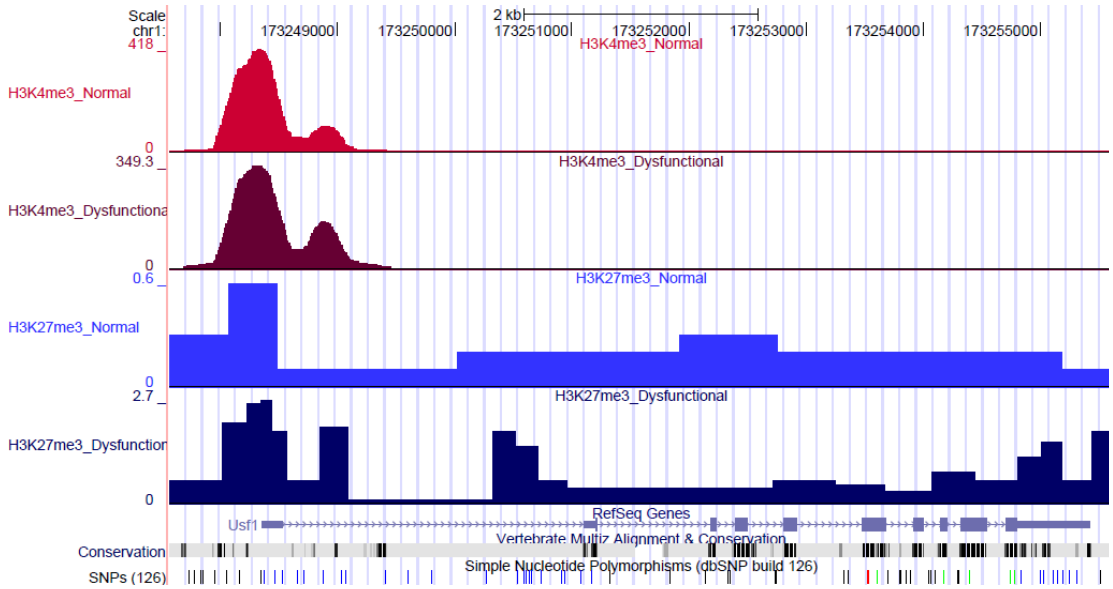


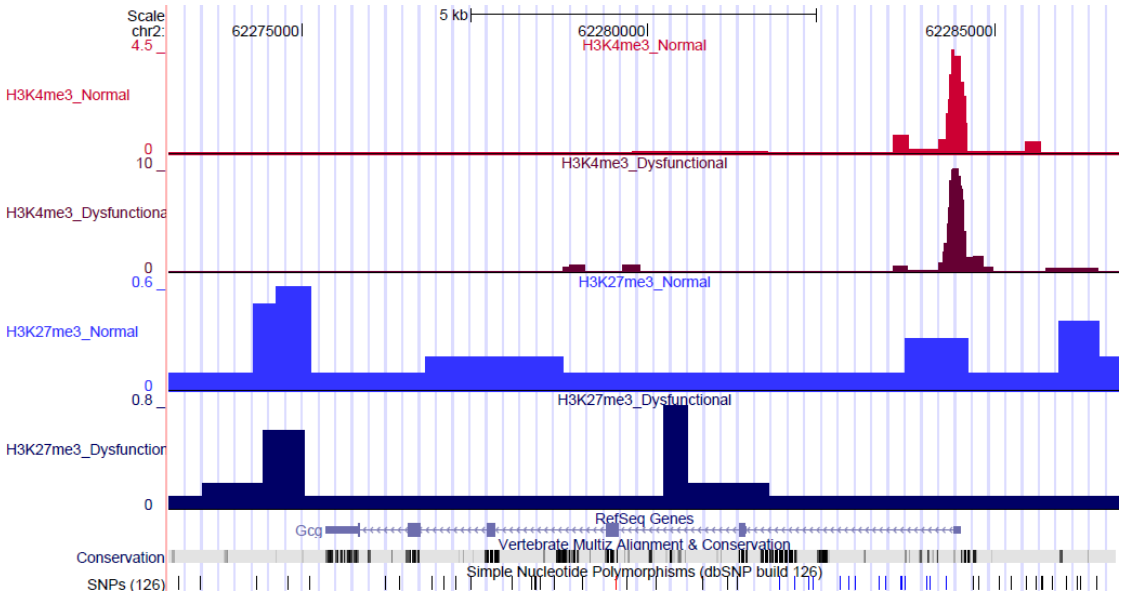
Figure 47. In dysfunctional β-cells, genes important for β-cell function are marked by decreased H3K4me3 levels and increased H3K27me3 levels at their promoter. H3K4me3 and H3K27me3 ChIP-Seq was performed in normal and dysfunctional NIT-1 cells and the presence or absence of these marks in important β-cell genes was visualized with the UCSC genome browser. (A) *Pdx1*, (B) *NeuroD1*, (C) *MafA*, (D) *Ins1*, (E) *Glut2*, (F) *Snap25*, (G) *Vamp2*, (H) *Cacna1a*, (I) *Chgb*, (J) *Kirrel6.2*, (K) *Sytl4*, (L) *Stxbp1*, and (M) *Syt7*, (N) *Ucp2* (O) *Usf1*.

Genes that are normally not expressed in β-cells include *Gcg* (an α-cell expressed gene) and *Sst* (secreted by δ-cells of the pancreas). There was increased H3K4me3 at those two loci in dysfunctional β-cells (Figure 48A, B). The T2D-associated genes *Kcnq1* (Figure 48C) and *Tcf7l2* (Figure 48D) showed an increase in K4me3 and a decrease in H3K27me3 in dysfunctional β-cells respectively. *Kcnq1*-over-expression impairs GSIS[285], and the over-expression of *Tcf7l2* exerted minor inhibitory effects on glucose-regulated changes in intracellular calcium concentrations and insulin release[286].

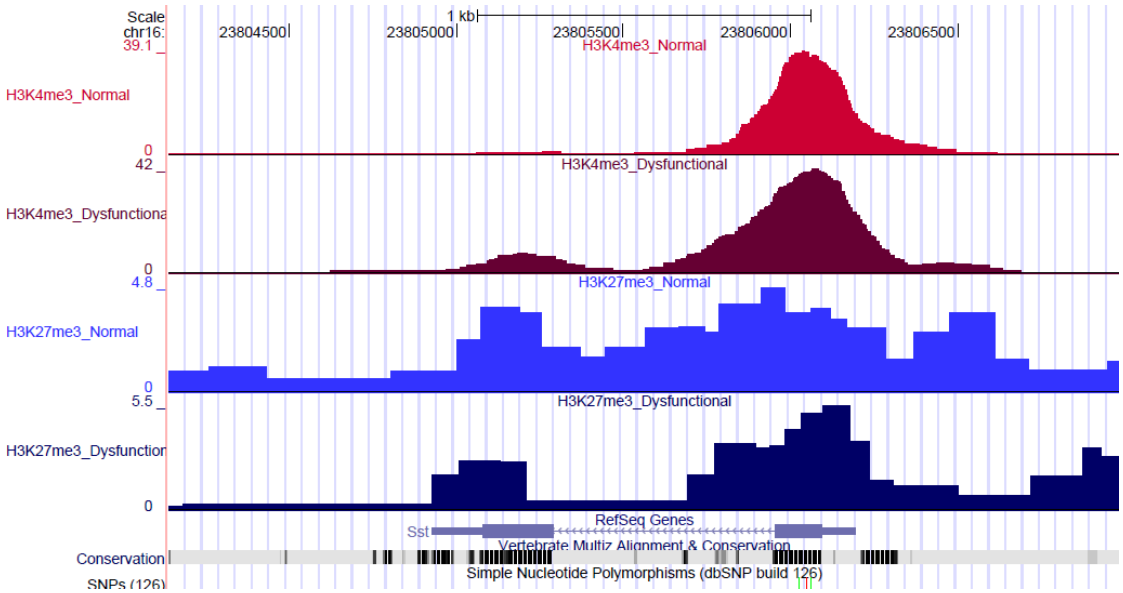
To see if a change in the chromatin state was associated with gene expression, I designed primers against these genes and looked at mRNA changes in normal and dysfunctional β-cells. There was a decrease in the expression for most of the genes which had lower H3K4me3 in β-cells, e.g. *Pdx1*, *NeuroD1*, *MafA*, *Ins1*, *Glut2*, *Snap25*, *Vamp2*, *Cacna1a*, *Chgb*, *Kirrel6.2*, *Sytl4*, *Stxbp1*, *Syt7* (Figure 49). There was a 3.5-fold increase in *Ucp2* mRNA levels too, consistent with the fact that hyperglycemia-induced mitochondrial superoxide production activates Ucp2, which decreases the ATP/ADP ratio and thus reduces the insulin-secretory response[262]. There was a 2-fold increase in the expression of *Gcg* and *Sst* (Figure 49) that were marked by increased H3K4me3 in dysfunctional β-cells. There was also a 2-fold increase in the T2DM-associated genes *Kcnq1* and *Tcf7l2* in dysfunctional β-cells (Figure 49).

To summarize, in dysfunctional β-cells, there was a decrease in H3K4me3 at important β-cell genes that correlated with a decrease in gene expression, and increase in H3K4me3 and the expression of genes not normally expressed in β-cells.

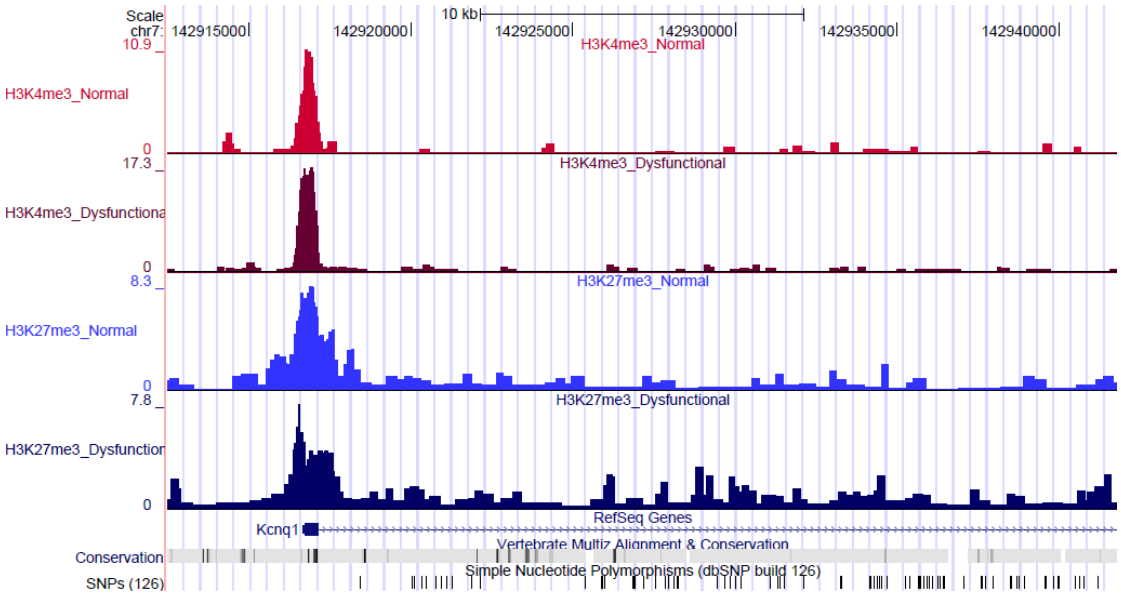
(A) Gcg


Scale
chr2:
62275000
5 kb
62280000
62285000
4.5
H3K4me3_Normal
H3K4me3_Normal
0
10
H3K4me3_Dysfunctional
H3K4me3_Dysfunctiona
0
0.6
H3K27me3_Normal
H3K27me3_Normal
0
0.8
H3K27me3_Dysfunctional
H3K27me3_Dysfunctior
0
RefSeq Genes
Gcg
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

(B) Sst


Scale
chr16:
23804500
1 kb
23805000
23805500
23806000
23806500
39.1
H3K4me3_Normal
H3K4me3_Normal
0
42
H3K4me3_Dysfunctional
H3K4me3_Dysfunctiona
0
4.8
H3K27me3_Normal
H3K27me3_Normal
0
5.5
H3K27me3_Dysfunctional
H3K27me3_Dysfunctior
0
RefSeq Genes
Sst
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

(C) Kcnq1



(D) Tcf7l2

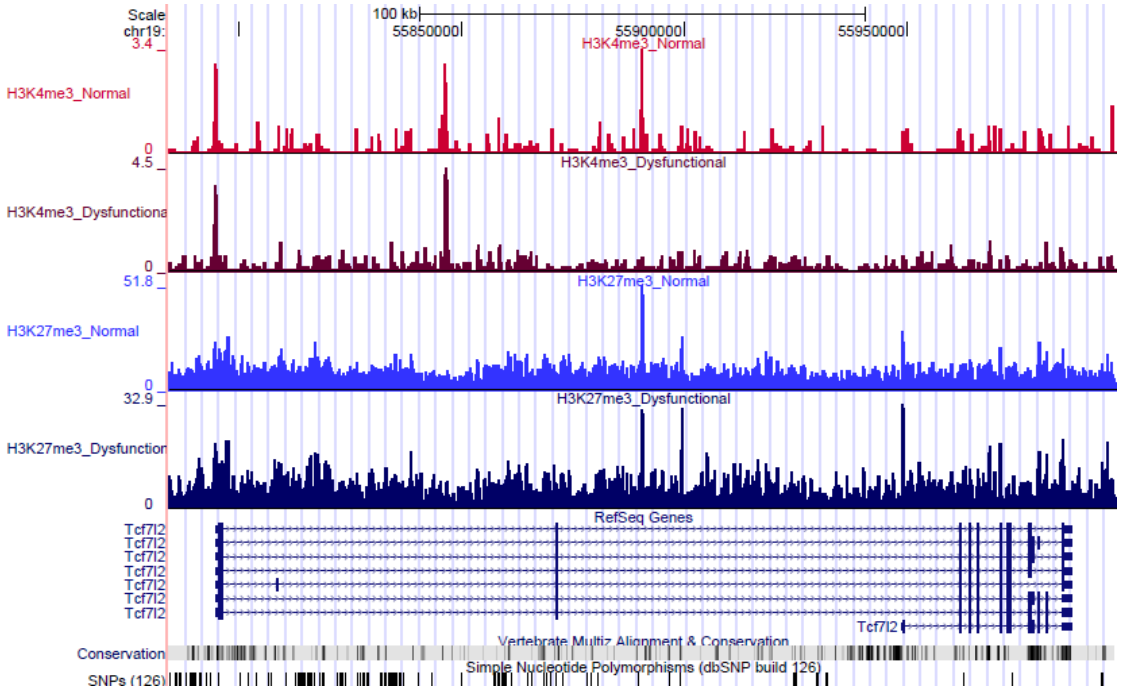


Figure 48. In dysfunctional β-cells, non- or lowly-expressed genes β-cell genes are marked by increased H3K4me3 levels and decreased H3K27me3 levels at their promoter. H3K4me3 and H3K27me3 ChIP-Seq was performed in normal and dysfunctional NIT-1 cells and the presence or absence of these marks in important β-cell genes was visualized with the UCSC genome browser. (A) *Gcg*, (B) *Sst*, (C) *Kcnq1*, (D) *Tcf7l2*.

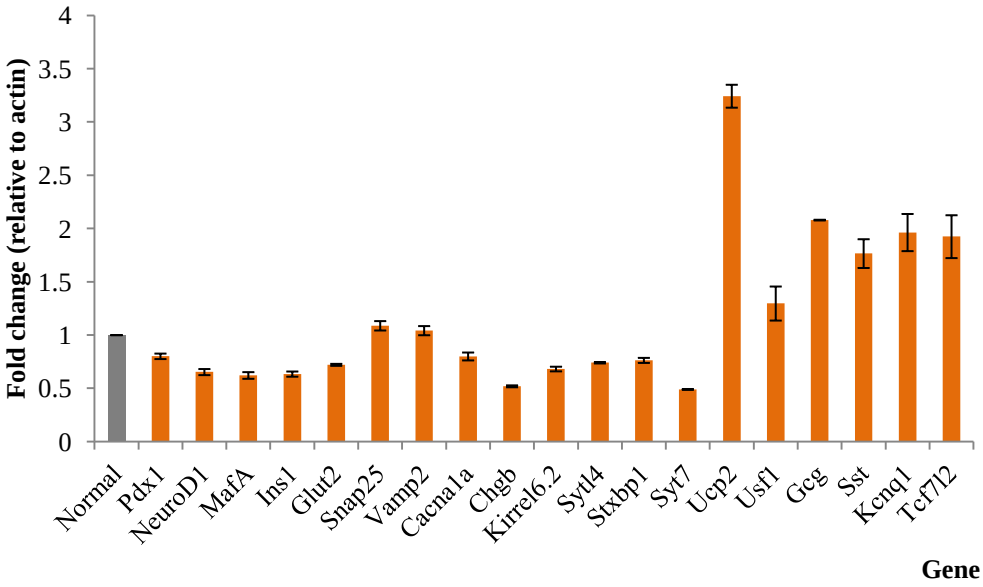


Figure 49. Genes important for normal β-cell function are expressed at lower levels in dysfunctional β-cells, while genes normally expressed at low levels in dysfunctional β-cells have increased gene expression. NIT-1 β-cells were treated with 20mM glucose and 0.4mM palmitate for 48hr and RNA harvested from untreated and treated cells (n=4 each). Transcript levels of these genes were assayed by RT-PCR, mean values ± S.E. are plotted as fold change relative to the level of the corresponding gene expression in the untreated control, normalized to endogenous *β-actin*.

5.2.5. Dynamic Association of Pdx1, NeuroD1 and MafA with Chromatin in Normal and Dysfunctional Beta-cells

There is a more open chromatin configuration in dysfunctional β-cells (Chapter 5.2.2.) and dysfunctional β-cells are more highly enriched in genes involved in general cellular regulatory activities and the specification of other developmental lineages (Chapter 5.2.3.). Dysfunctional β-cells also have reduced levels of H3K4me3 at transcription factors and other genes required for proper β-cell function, and increased H3K4me3 levels in some genes not normally expressed in β-cells (Chapter 5.2.4.), which mostly correlated with corresponding gene expression changes. In glucolipotoxic conditions, *MafA* expression is inhibited, whereas Pdx1 is affected at the post-translational level in its ability to translocate to the nucleus[280]. It is not known what effects high glucose and lipids have on the expression of *NeuroD1*, but Figure 51 shows that mRNA levels of *NeuroD1* are reduced in the dysfunctional β-cells I generated, along with that of *Pdx1* and *MafA*. I therefore hypothesized that chromatin remodelling in dysfunctional β-cells is in part brought about by the loss of key β-cell transcription factors, and that changes in that would affect chromatin states.

To look at transcription factor association with H3K4me3 and H3K27me3 on a genome-wide scale, the surrounding regions (±5kb) of each TFBS were examined for the tag distribution of H3K4me3 and H3K27me3. For the co-occupancy of the transcription factors, the TFBSs that overlapped within 50bp were merged into a single peak. The numbers of H3K4me3 and H3K27me3 tags were counted according to the distance from the center of the TFBS and log10 transformed for modification level. The minimum modification level for H3K4me3 and H3K27me3 among the three TFBSs and co-occupied sites was taken and used as the basal level. Each modification level was normalized based on the basal level of H3K4me3 and H3K27me3 respectively. As Figure 50 shows, the Pdx1 (green), NeuroD1 (blue), MafA (red) are all highly marked by H3K4me3 in normal β-cells when compared to the low level of Pdx1 (light green), NeuroD1 (light blue), MafA (yellow) TFBS marked with H3K27me3,

demonstrating that there is an open state of chromatin in normal β-cells when Pdx1, NeuroD1 and MafA are bound.

To explore this in greater detail, TFBS (±65bp from peak center) of either Pdx1, NeuroD1 and MafA were overlapped with promoters of genes (± 2kb from TSS) that are switched on or off in NIT-1 cells. This was to look at all the genes that are expressed in normal β-cells and seeing whether they would be switched on when either Pdx1, NeuroD1 or MafA are bound. There are a lot fewer genes that are switched off (15%) when any of these three transcription factors are bound, compared to the 45% genes that are switched off when none of them are bound (Figure 51). This indicates that transcription factor binding has an activating effect on gene expression.

Interestingly, when the TFBS were overlapped with H3K4me3 and H3K27me3 in dysfunctional β-cells, there was a 2-fold increase in the levels of Pdx1 (light green), NeuroD1 (light blue), MafA (yellow) and triply occupied (grey) TFBS associated with H3K27me3 while levels of Pdx1 (green), NeuroD1 (blue), MafA (red) and triply occupied (black) TFBS associated with H3K4me3 remained at similar levels as those in normal β-cells, p values < 1e-100 (Figure 52). To look at this from a different angle, Figure 52C shows that TFBS overlapped with H3K4me3 in normal and dysfunctional β-cells have similar levels of H3K4me3 modification, and Figure 52D shows that TFBS overlapped with H3K27me3 in normal and dysfunctional β-cells has a 2-fold higher level of H3K27me3 modification in dysfunctional β-cells.

I hypothesized that the increase in the repressed chromatin state in dysfunctional β-cells that were associated with any of the three transcription factors would also have a decrease in gene expression, as H3K27me3 is an indicator of a repressive chromatin state. To this end, I ran a microarray on RNA extracted from normal and dysfunctional β-cells (n=4 each). Figure 53A shows that genes which had a TFBS and were marked by a increase in H3K27me3 in

dysfunctional β-cells also had a decrease in gene expression. This data is shown from another angle in Figure 53B, where dysfunctional β-cells show a higher H3K27me3 at the Pdx1 TFBS (p value = 3.7E-48), NeuroD1 TFBS (p value = 5.8E-85) and MafA TFBS (p value = 3.4E-85). A full list of genes from the microarray is listed in Supplementary Table 9.

This clearly indicates that the regions bound by these transcription factors in normal β-cells tend to have an increased level of H3K27me3 in dysfunctional β-cells, and this is associated with a decrease in gene expression. This is in contrast to the overall more open state of chromatin in dysfunctional β-cells, as shown by the increased levels of H3K4me3 and FAIRE at promoter regions and decreased levels of H3K27me3 at promoter regions (Chapter 5.2.2-4).

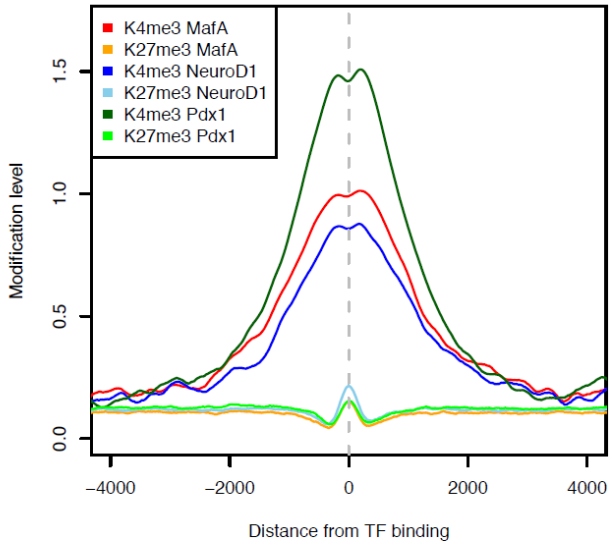


Figure 50. Pdx1, NeuroD1 and MafA TFBS are enriched for H3K4me3 in normal β-cells. The surrounding regions (±5kb) of each TFBS were examined for the tag distribution of H3K4me3 and H3K27me3. The numbers of H3K4me3 and H3K27me3 tags were counted according to the distance from the center of the TFBS and then log10 transformed for modification level. The minimum modification level for H3K4me3 and H3K27me3 among the three TFBSs was used as the basal level. Each modification level was normalized based on the basal level of H3K4me3 and H3K27me3 respectively. (Dr Choi performed this analysis.)

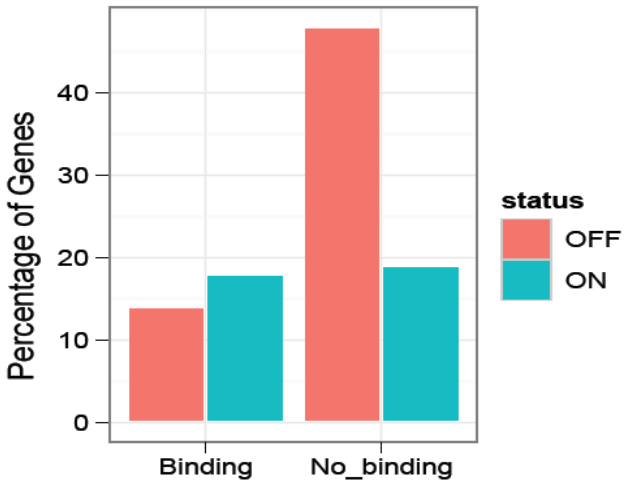


Figure 51. Transcription factor binding has an activating effect on gene expression in normal β-cells. TFBS (±65bp from peak center) of either Pdx1, NeuroD1 and MafA were overlapped with promoters of genes (± 2kb from TSS) that are switched on or off in NIT-1 cells. (Gireesh Bogu performed this analysis.)

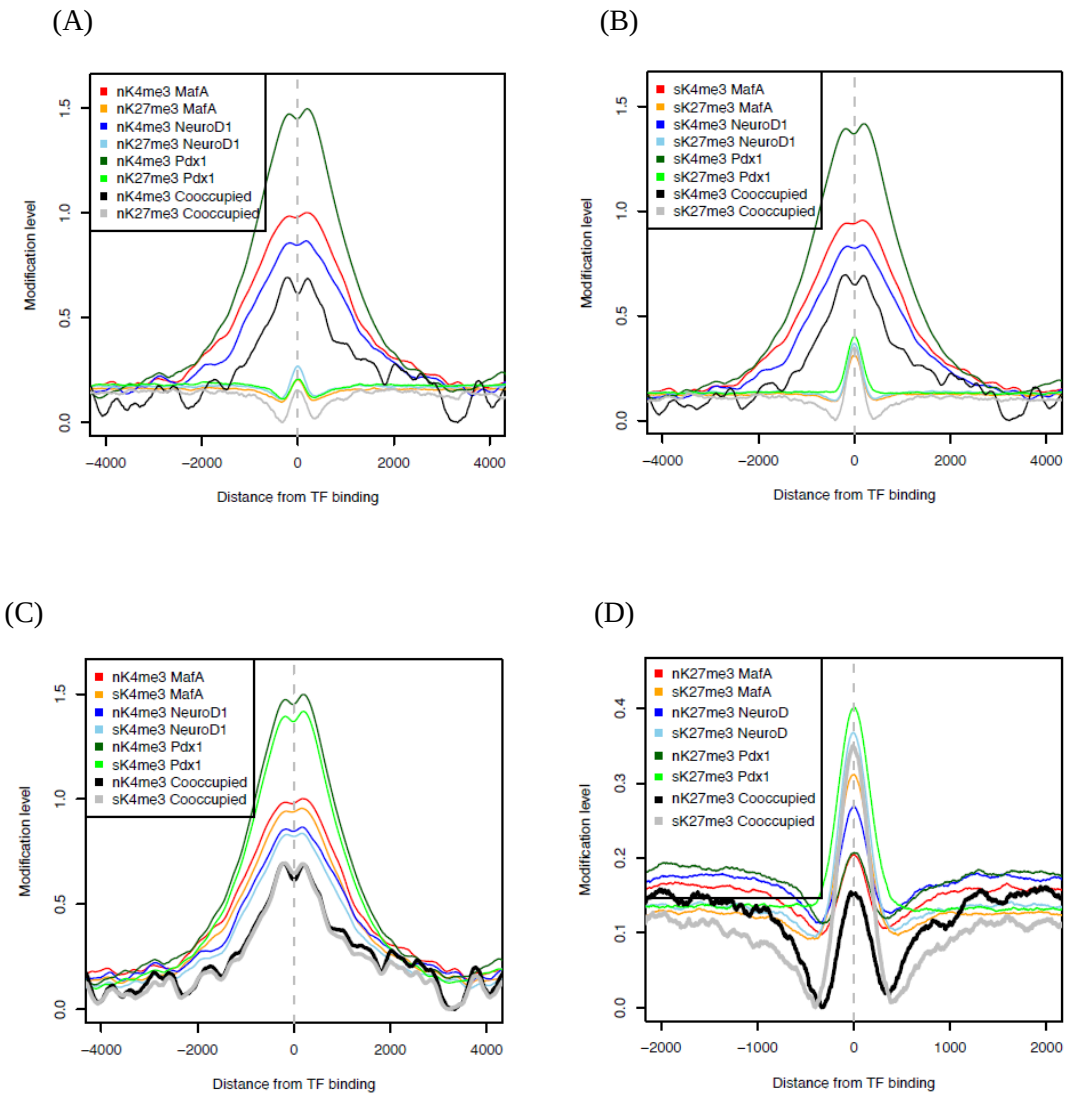
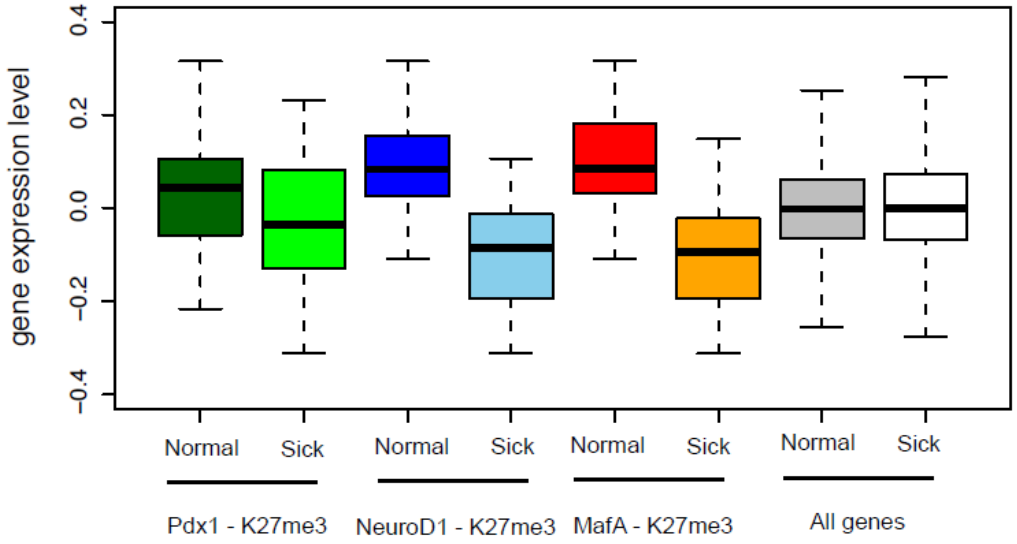


Figure 52. Pdx1, NeuroD1 and MafA TFBS are enriched for H3K4me3 in normal β-cells and display increased levels of H3K27me3 in dysfunctional β-cells. TFBS were overlapped with H3K4me3 and H3K27me3 in dysfunctional β-cells and there was a 2-fold increase in the levels of Pdx1 (light green), NeuroD1 (light blue), MafA (yellow) and triply occupied (grey) TFBS associated with H3K27me3 (B) while levels of Pdx1 (green), NeuroD1 (blue), MafA (red) and triply occupied (black) TFBS associated with H3K4me3 remained at similar levels as those in normal β-cells (A). To look at this from a different angle, (C) shows that TFBS overlapped with H3K4me3 in normal and dysfunctional β-cells have similar levels of H3K4me3 modification, and (D) shows that TFBS overlapped with H3K27me3 in normal and dysfunctional β-cells had a 2-fold higher level of H3K27me3 modification in dysfunctional β-cells (p values < 1E-100). N=normal β-cells, S=dysfunctional β-cells. (Dr Choi performed this analysis.)

(A)



(B)

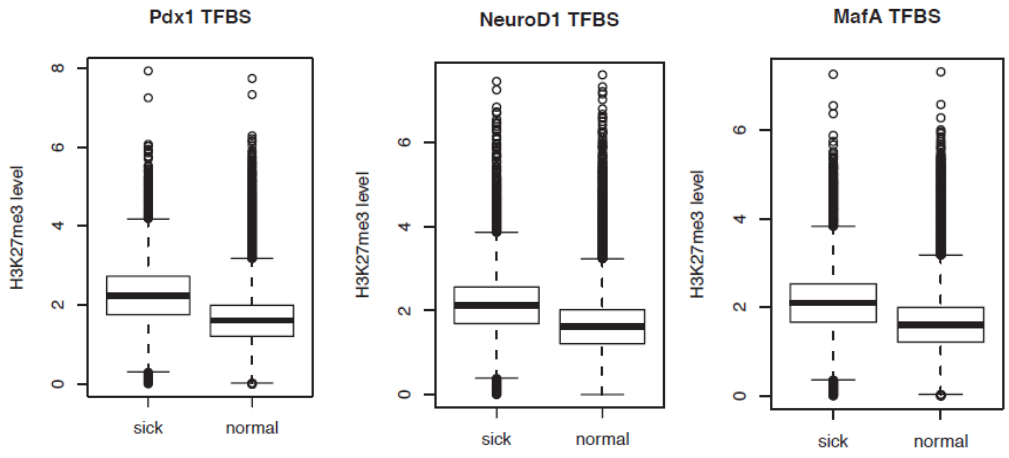


Figure 53. (A) Genes bound by either Pdx1, NeuroD1 or MafA that were marked by a increase in H3K27me3 in dysfunctional β-cells also had a decrease in gene expression. (B) H3K27me3 levels at the TFBS in the dysfunctional (sick) and normal β-cells. Dysfunctional β-cells show a higher H3K27me3 at the Pdx1 TFBS (p value 3.7E-48), NeuroD1 TFBS (p value p = 5.8E-85) and MafA TFBS (p value = 3.4E-85). Sick=dysfunctional β-cells. (Dr Choi performed this analysis.)

I wanted to look in greater detail at the network of genes bound by each transcription factor that was associated with H3K27me3 in dysfunctional β-cells. The transcription factor ChIP-Seq peaks that overlapped with H3K27me3-ChIP-Seq in dysfunctional β-cells were imported into IPA. Networks associated with its biological, molecular and cellular function, physiological system development and function were generated.

Genes bound by Pdx1 in normal β-cells and showing an increase in H3K27me3 in dysfunctional β-cells were enriched in amino acid and carbohydrate metabolism, cell cycle, digestive system development and cellular activities such as molecular transport, RNA trafficking, damage and repair (Table 12). Genes bound by NeuroD1 in normal β-cells and showed an increase in H3K27me3 in dysfunctional β-cells were enriched in general cellular activities, endocrine development and function, embryonic and hematological development, and amino acid metabolism (Table 13). Genes bound by MafA in normal β-cells and showed an increase in H3K27me3 in dysfunctional β-cells were enriched in molecular transport, DNA replication, RNA post-transcriptional modification, gastrointestinal disease, cellular and embryonic development, cell signaling, cell cycle and cell death (Table 14). A full list of genes involved in each of those networks, pathways and functions is listed in Supplementary Table 10.

Taking into account there was reduced expression of Pdx1, NeuroD1 and MafA in dysfunctional β-cells, I have shown that the H3K27me3-marked genes (that have a TFBS) in dysfunctional β-cells are enriched for genes involved in metabolic and general cellular regulatory processes.

Associated Network Functions	Network Score
Amino Acid Metabolism, Drug Metabolism, Molecular Transport	34
Carbohydrate Metabolism, Small Molecule Biochemistry, Cell Cycle	34
Cell Cycle, Organismal Functions, Cellular Assembly and Organization	31
Behavior, Digestive System Development and Function, Cell Signaling	31
Molecular Transport, RNA Trafficking, RNA Damage and Repair	31

Diseases and Disorders	p-value	# Molecules
Endocrine System Disorders	1.88E-34	711
Gastrointestinal Disease	1.88E-34	953
Genetic Disorder	1.88E-34	1730
Metabolic Disease	1.88E-34	801
Neurological Disease	5.17E-23	1056

Molecular and Cellular Functions	p-value	# Molecules
Gene Expression	1.07E-16	565
Cell Cycle	1.20E-15	407
Cell Death	2.28E-13	763
Cellular Growth and Proliferation	1.30E-10	809
Cellular Development	1.24E-08	671

Physiological System Development and Function	p-value	# Molecules
Organismal Development	1.59E-12	340
Organismal Survival	4.33E-12	332
Embryonic Development	2.60E-10	249
Nervous System Development and Function	1.12E-07	497
Organ Development	8.47E-07	288

Table 12 Top biological networks and pathways associated with Pdx1 binding sites that are marked by H3K27me3 in dysfunctional β-cells.

Pdx1 ChIP-Seq peaks were identified with MACS and overlapped with H3K27me3-ChIP-Seq in dysfunctional β-cells, and IPA was used to generate this list of networks associated with its biological, molecular and cellular function, physiological system development and function. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible molecules in the dataset.

Associated Network Functions	Network Score
Cell Morphology, Cellular Growth and Proliferation, Hematological System Development and Function	39
Endocrine System Development and Function, Cell Morphology, Small Molecule Biochemistry	36
Cellular Development, Embryonic Development, Tissue Development	34
Hematological System Development and Function, Hematopoiesis, Tissue Morphology	34
Amino Acid Metabolism, Small Molecule Biochemistry, DNA Replication, Recombination, and Repair	34

Diseases and Disorders	p-value	# Molecules
Genetic Disorder	2.62E-54	1405
Endocrine System Disorders	1.49E-48	609
Gastrointestinal Disease	1.49E-48	850
Metabolic Disease	1.49E-48	667
Neurological Disease	6.50E-44	894

Molecular and Cellular Functions	p-value	# Molecules
Cell Death	9.60E-12	578
Cellular Development	2.36E-11	517
Cellular Assembly and Organization	2.41E-10	264
Cell Morphology	1.05E-09	293
Cellular Growth and Proliferation	2.30E-09	595

Physiological System Development and Function	p-value	# Molecules
Tissue Development	1.87E-10	341
Organismal Survival	1.18E-09	209
Nervous System Development and Function	1.84E-09	403
Behavior	2.27E-08	168
Organ Development	2.35E-08	211

Table 13. Top biological networks and pathways associated with NeuroD1 binding sites that are marked by H3K27me3 in dysfunctional β-cells.

NeuroD1 ChIP-Seq peaks were identified with MACS and overlapped with H3K27me3-ChIP-Seq in dysfunctional β-cells, and IPA was used to generate this list of networks associated with its biological, molecular and cellular function, physiological system development and function. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible molecules in the dataset.

Associated Network Functions	Network Score
Molecular Transport, Protein Trafficking, DNA Replication, Recombination, and Repair	34
RNA Post-Transcriptional Modification, Gastrointestinal Disease, Genetic Disorder	34
Cellular Development, Embryonic Development, Tissue Development	32
Cell Signaling, Auditory Disease, Dermatological Diseases and Conditions	32
Cell Cycle, DNA Replication, Recombination, and Repair, Cell Death	32

Diseases and Disorders	p-value	# Molecules
Genetic Disorder	1.38E-59	1819
Endocrine System Disorders	3.22E-50	759
Gastrointestinal Disease	3.22E-50	1092
Metabolic Disease	3.22E-50	846
Neurological Disease	2.18E-46	1137

Molecular and Cellular Functions	p-value	# Molecules
Cell Death	3.58E-13	760
Gene Expression	6.65E-13	538
Cellular Development	1.02E-10	660
Cellular Assembly and Organization	5.24E-10	357
Cellular Growth and Proliferation	8.78E-10	777

Physiological System Development and Function	p-value	# Molecules
Cell-mediated Immune Response	5.02E-07	161
Hematological System Development and Function	5.02E-07	248
Hematopoiesis	6.02E-07	217
Nervous System Development and Function	1.08E-06	485
Behavior	1.25E-06	202

Table 14. Top biological networks and pathways associated with MafA binding sites that are marked by H3K27me3 in dysfunctional β-cells.

MafA ChIP-Seq peaks were identified with MACS and overlapped with H3K27me3-ChIP-Seq in dysfunctional β-cells, and IPA was used to generate this list of networks associated with its biological, molecular and cellular function, physiological system development and function. The score is a numerical value used to rank networks according to their degree of relevance to the network eligible molecules in the dataset.

5.2.6. BIVALENCY IN THE NORMAL AND DYSFUNCTIONAL BETA-CELL

Bivalent chromatin contains both activating and repressive marks in the same area. Activating chromatin modifications increase the accessibility of the chromatin to RNA polymerase, where repressing modifications decrease the accessibility to RNA polymerase. Usually, these do not both occur at the same location, as it seems they would have opposing effects; however, they are both present in bivalent chromatin. The repressing modifications usually take precedence, causing the gene to become inactivated. Once the repressing modifications are removed however, the activating modifications attract transcription machinery, and the gene becomes activated. Bivalent chromatin was first identified in embryonic stem cells and they were proposed to serve a role in genetic imprinting and in development[137]. Bivalent domains are also found in other cell types, e.g. mouse embryonic fibroblasts, neural progenitors[138] and human T cells[139]. A recent comparison of H3K4me3 and H3K27me3 in islets show that these bivalent domains are usually mutually exclusive or absent, but there are few genes that had high levels of both marks, e.g. the *Hox* gene clusters and neuronal transcription factors, perhaps indicating a retention of "memory" of earlier transcription or the re-activation of these genes under certain metabolic conditions[140]. Therefore I hypothesized that bivalent domains exist in β-cells and not only in islets that comprise of other endocrine cell types, and that some of these bivalent domains would change in dysfunctional β-cells, possibly affecting its impaired ability to secrete insulin in response to a glucose stimulus.

There were 5288 bivalent genes in normal β-cells, and of those, 693 genes lost the H3K27me3 mark in dysfunctional β-cells, while 38 genes lost the H3K4me3 mark in dysfunctional β-cells. A full list of bivalent genes thus marked is available in Supplementary Tables 11A-C. This is consistent with the fact that there was an increase in the numbers of H3K4me3 and FAIRE peaks, and decrease in the global numbers of H3K27me3 peaks in dysfunctional β-cells. As there was such a high number of bivalent genes in normal β-cells that lost the

H3K27me3 mark in dysfunctional β-cells compared to those that lost the H3K4me3 mark in dysfunctional β-cells, I wanted to know how similar or different these genes were in terms of their biological function, when compared to the bivalent genes identified in normal β-cells. The RefSeq identity list of nearest associated bivalent genes indentified in normal β-cells, together with the nearest associated bivalent genes in normal β-cells that had lost H3K27me3 in dysfunctional β-cells, was uploaded into IPA[187]. The top biological networks and cellular processes and functions associated with them were identified, and I utilized the comparison function in IPA to compare these two datasets.

Strikingly, there was a much higher over-enrichment for bivalent genes in normal β-cells that were involved in many aspects of developmental processes when compared to bivalent genes in normal cells that had lost the H3K27me3 in dysfunctional β-cells, e.g. organ, cellular, tissue, embryonic, nervous system, connective, skeletal development (Figure 54). There was also a higher enrichment for genes involved in disorders in the same list, e.g. genetic, neurological, skeletal and muscular, cardiovascular, connective tissue, developmental, inflammatory, gastrointestinal, immunological, physiological, and metabolic diseases. A full list of genes involved in these pathways is listed in Supplementary Table 12. I did not compare the 38 bivalent genes in normal β-cells that had lost the H3K4me3 mark in dysfunctional β-cells because there were too few genes in that list to yield a statistically significant result.

Taken together, this suggests that genes that are bivalently marked in normal β-cells are more involved in developmental processes and cell regulatory processes, compared to genes that are bivalently marked in normal β-cells and have lost the repressive H3K27me3 mark in dysfunctional β-cells.

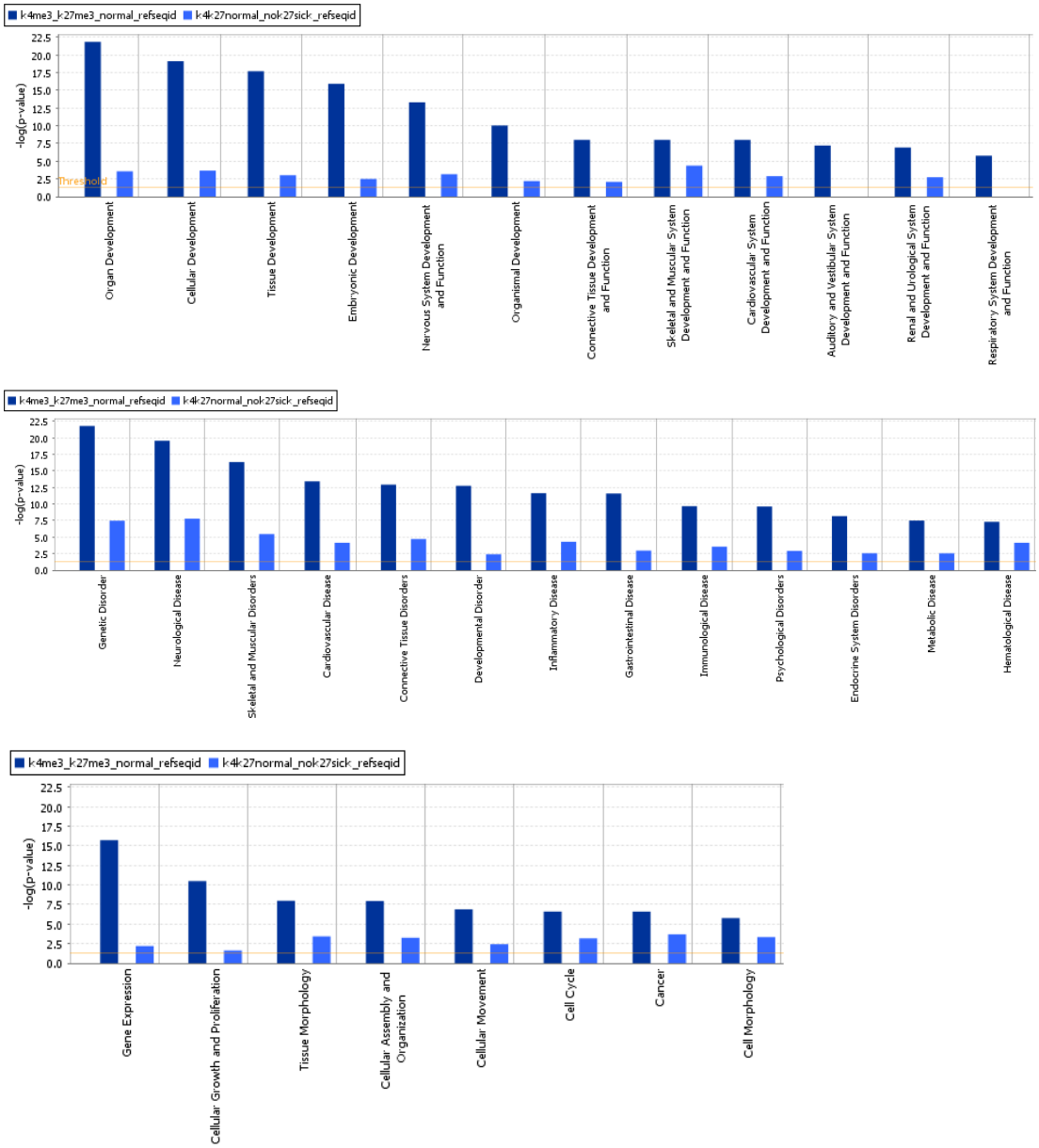


Figure 54. Genes that are bivalently marked in normal β-cells are more involved in developmental processes and cell regulatory processes, compared to genes that are bivalently marked in normal β-cells and have lost the repressive H3K27me3 mark in dysfunctional β-cells.

Normal=normal β-cells, S=dysfunctional β-cells.

For a more rigorous analysis of the H3K4me3 and H3K27me3 overlaps, the H3K4me3 and H3K27me3 peaks were overlapped and the genes containing those peaks were identified. This is in contrast to the above-mentioned overlaps of genes containing both H3K4me3 and H3K27me3, as that did not guarantee that the peaks were actually overlapping, just that the genes had to contain both H3K4me3 and H3K27me3 peaks. With the more stringent method of analysis, there were 113 bivalent genes in normal β-cells, with 16 genes losing the H3K4me3 mark (i.e. 16 genes with lower H3K4me3 mark in dysfunctional β-cells) and 98 genes losing the H3K27me3 mark in dysfunctional β-cells. A full list of bivalent genes with peak overlaps is available in Supplementary Table 11D. This trend is similar to what is observed when the less stringent H3K4me3 and H3K27me3 overlaps were used.

The developmental *Hox* gene cluster was bivalently marked (Figure 55), as identified by a group who profiled H3K4me3 and H3K27me3 in human pancreatic islets[140]. H3K4me3 levels in the *Hox* gene cluster were slightly increased and H3K27me3 levels slightly decreased in dysfunctional β-cells when compared to normal β-cells. Bivalent genes in normal β-cells that lost H3K27me3 in dysfunctional β-cells include those involved in pancreatic development and/or function, e.g *MafB, Oxidation resistance (Oxr)-1, Calcium/calmodulin-dependent protein kinase (Camk)-2b, miR-9, SRY (sex determining region Y)-box (Sox)-17* (Figure 56A-E); genes involved in neuronal development and function, e.g. *Semaphorin (Sema)-3d* and *Cadherin EGF LAG seven-pass G-type receptor (Celsr)-1* (Figure 56F, G); and members of the *Fox* family transcription factors e.g. *Foxq1* and *Foxf2* (Figure 56H, I). The potassium channel *Potassium voltage-gated channel Shaw-related subfamily member (Kcnc)-1* is also bivalently marked in β-cells (Figure 56J). Of great interest is the T2DM-identified potassium channel *Kcnq1* (Figure 58A).

There was an increase in the levels of H3K4me3 and decrease in H3K27me3 at the *MafB* gene in dysfunctional β-cells (Figure 56A), however this did not correlate with levels of gene expression, which remained the same as in normal β-cells (Figure 57). In embryos, *MafB* is

expressed before *MafA*, and the differentiation of β-cells proceeds through a *MafB*$^{+}$ *MafA*$^{-}$ *Ins*+ intermediate cell to *MafB*$^{-}$ *MafA*$^{+}$ *Ins*$^{+}$ cells[70]. Furthermore, the MafB to MafA transition follows induction of *Pdx-1* expression in *MafB*$^{+}$ *Ins*+ cells, suggesting that MafB may have a dual role in regulating embryonic differentiation of both α- and β-cells while MafA regulates replication/survival and function of β-cells after birth[70]. Within the adult pancreas, *MafB* is only expressed in islet α-cells and contributes to cell type–specific expression of the *Gcg* gene[287]. The fact that the chromatin environment around the *MafB* gene in normal β-cells was bivalently marked in normal β-cells, and that there was a shift towards a more permissive and open chromatin state in dysfunctional β-cells is extremely interesting. Perhaps this is the chromatin signature of a gene that was required for earlier embryonic development and not in the mature β-cell, but it still retained the epigenetic memory of its earlier expression, and is poised to "revert" back to its previous cellular state should an appropriate stimulus occur. However, it is important to note that this shift towards a more open chromatin state did not lead to an increase in *MafB* gene expression (despite there being a 2-fold increase in the glucagon gene expression in dysfunctional β-cells (Figure 49), suggesting that glucolipotoxicity is merely triggering a change in chromatin environment at the *MafB* gene, perhaps allowing for the possibility for transcriptional activators to bind and mediate changes in gene expression, should additional/ prolonged stimulus occur.

There was an increase in H3K4me3 and decrease in H3K27me3 at the *Oxr1* gene in dysfunctional β-cells (Figure 56B), and this was coupled with a 2.2-fold increase in gene expression (Figure 57). Oxr1 is the receptor for Orexin A (Oxa), and Oxa plays a critical role in the regulation of sleep/wake states, feeding behaviour, reward processes and the regulation of food intake in a dose-dependent fashion[288]. Oxa increases food intake in rodents, and fasting activates Oxa neurons in both the lateral hypothalamic area and gut[288]. Oxa is also present in the pancreas, and depolarizing stimuli evoke the release of Oxa from rat pancreatic islets in a calcium-dependent manner, causing an increase in glucagon secretion and reducing insulin secretion[289]. Oxr1 co-localizes with insulin β-cells and glucagon α-cells in normal rats,

and the incidence of *Oxr1*-expressing cells in diabetic rats treated with streptozotocin was significantly increased, along with an increased co-localization with glucagon-expressing cells[290]. In the context of Oxr1 being bivalently marked in normal β-cells and *having* a more open chromatin state in dysfunctional β-cells that correlated with an increase in Oxr1 mRNA levels, and taking into account *Gcg* gene expression increases 2-fold in dysfunctional β-cells while *Ins1* expression decreases 40% (Figure 49), this suggests that the mechanism of chromatin opening at the *Oxr1* gene could be in part responsible for the decrease in insulin secretion and the increase in *Gcg* expression in diabetic rats.

There was an increase in H3K4me3 in dysfunctional β-cells and concomitant decrease in H3K27me3 at the *Camk2b* gene (Figure 56C), and there was also a 2-fold increase in mRNA levels of *Camk2b* (Figure 57). Camk2b activity is required to induce *MafA* expression under high glucose conditions[291], and this is consistent with the fact that dysfunctional β-cells were exposed to a higher glucose concentration (20mM) than normal β-cells (7mM). *MafA* mRNA expression was reduced in dysfunctional β-cells (Figure 49), indicating that this modest increase in *Camk2b* expression was not sufficient to combat the overall effects that glucolipotoxicity had on *MafA* expression.

miR-9 displays higher H3K4me3 in dysfunctional β-cells (Figure 56D). This is consistent with the fact that miR-9 reduces exocytosis in GSIS via diminishing the expression of *Onecut-2*, which in turn increases the level of *Granuphilin*[292]. Granuphilin is a Rab-GTPase effector associated with β-cell secretory granules and negatively affects insulin release. miR-9 also targets and reduces *Sirtuin (Sirt)-1* expression in insulin-secreting cells[293]. Sirt1 positively regulates insulin secretion in pancreatic β-cells by repressing the *Ucp2* gene[294]. Therefore it is possible that mir-9 in dysfunctional β-cells acts to reduce exocytosis during GSIS via similar mechanisms.

Sox17 is also bivalently marked in normal β-cells (Figure 56E), with low levels of H3K4me3 and higher levels of H3K27me3. There is a very slight decrease in levels of H3K27me3 in dysfunctional β-cells, but that did not correlate with changes in gene expression (Figure 57). Sox17 is a transcription factor required for the normal development of the definitive gut endoderm [295] and physically interacts with β-catenin (a component of the Wnt-signalling pathway) in order to potentiate its transcriptional activation of its target endodermal genes[296]. It is uncertain whether it is expressed in the adult mouse pancreas[297,298], and is very lowly expressed in NIT-1 normal and dysfunctional β-cells. Sox17 is bivalently marked in human ESCs and upon hESC differentiation, expression of *Sox17* in definitive endoderm cells appears to be activated by demethylation of histone H3K27, while suppression of *Sox17* in hepatocytes seems to be modulated by the methylation of H3K27[299]. Perhaps bivalency of Sox17 in β-cells is a remnant of its earlier bivalent state in ESCS, and though there was no noticeable change in the chromatin environment around *Sox17* in dysfunctional β-cells, perhaps additional/ prolonged stimulus is needed for an appreciable change at the *Sox17* gene locus.

There is an increase in H3K4me3 at the *Sema3d* gene in dysfunctional β-cells and a decrease in H3K27me3 (Figure 56F), which was accompanied by a 2.4-fold increase in gene expression (Figure 57). Sema3d induces the collapse and paralysis of neuronal growth cones, and could potentially act as repulsive cues toward specific neuronal populations[300]. Another neural gene *Celsr1* was also bivalently marked in normal β-cells and had an increase in H3K4me3 and decrease in H3K27me3 in dysfunctional β-cells (Figure 56G). However, this increase in open chromatin configuration resulted in a 40% decrease in *Celsr1* gene expression in dysfunctional β-cells (Figure 57). *Celsr1* is a neural gene involved in cell signalling during nervous system formation[301]. Taken together, it appears that increase in H3K4me3 in dysfunctional β-cells occurs at some neural genes that are normally bivalently marked.

There was also an increase in H3K4me3 in *Foxq1* and a slight decrease in H3kK27me3 in dysfunctional β-cells (Figure 56H), this was accompanied by a 1.8-fold increase in *Foxq1* mRNA (Figure 57). Like other members of the Fox family that are involved in embryonic development, cell cycle regulation, tissue-specific gene expression, cell signaling, and tumorigenesis, *Foxq1* is expressed in the visceral endoderm during development[302] and is over-expressed in colorectal cancers and enhances tumorigenicity and tumor growth presumably through its angiogenic and antiapoptotic effects[303]. *Foxq1* is also over-expressed in pancreatic cancer[304]. *Foxf2* was also bivalent marked in normal β-cells and had an increased H3K4me3 in dysfunctional β-cells (Figure 56I). However, there was an increase in the levels of H3K27me3 in dysfunctional β-cells too, with slight increases in gene expression (Figure 57). *Foxf2* is expressed during development in the mesodermal tissues involved in epithelio-mesenchymal interactions, and in the central nervous system, eye, ear, and limb buds[305]. Taken together, it suggests that genes bivalent marked in normal β-cells that are involved in specifying other developmental lineages have a more open chromatin configuration in β-cell dysfunction.

Kcnc1 is bivalently marked (Figure 56J) in normal β-cells, with a decrease in H3K4me3 and increase in H3K27me3 in dysfunctional β-cells, which correlates with a 55% decrease in gene expression (Figure 57). Kcnc1 mediates the voltage-dependent K^+ permeability of excitable membranes[306]. This suggests β-cell dysfunction leads to a more repressive chromatin environment around the Kcnc1 gene, resulting in a decrease in gene expression.

Kcnq1 is bivalently marked at its promoter (Figure 58A), and there is an increase in H3K4me3, which is accompanied by a 2-fold increase in *Kcnq1* mRNA expression levels (Figure 57). Kcnq1 is a voltage-gated potassium channel and is involved in repolarizing cardiac tissue following an action potential, and transporting water and salts in tissues[307]. Variants in KCNQ1 are associated with susceptibility to T2DM in the Korean, Chinese and European population[308] and variants in KCNQ1 predicts future T2DM in Scandinavian

populations, possibly by mediating impaired insulin secretion[309]. When *Kcnq1* was over-expressed in mouse insulinoma MIN-6 cells, both the density of the Kcnq1 current and the density of the total K^+ current were significantly increased. In addition, insulin secretion by glucose, pyruvate, or tolbutamide was significantly impaired, suggesting that increased Kcnq1 protein expression limits insulin secretion from pancreatic β-cells by regulating the potassium channel current[285]. Taken together, it appears *Kcnq1* is bivalently marked in normal β-cells and changes in the chromatin environment around its promoter leads to an increase in gene expression, thereby contributing to reduce GSIS in dysfunctional β-cells.

Very interestingly, KCNQ1 is located in an imprinted gene cluster at the telomeric end of mouse chromosome 7 which contains a differentially methylated CpG island, K_vDMR. K_vDMR is required for the imprinting of multiple genes, including the genes encoding the maternally expressed placental-specific transcription factor *Ascl2*, the cyclin-dependent kinase *Cdk1n1c*, and *Kcnq1*. K_vDMR maps within the intron of *Kcnq1* and contains the promoter for a paternally expressed 91kb non-coding (nc) antisense transcript *Kcnq1ot*, which plays a direct role in silencing the multiple mRNAs genes in the *Kcnq1*-imprinted gene cluster[310]. *Kcnq1ot* interacts with the Polycomb repressive complex 2 and with the H3K9-specific methyltransferase G9a, and its expression coincides with the assembly of repressive chromatin defined by H3K27me3 and H3K9me3[311].

Kcnq1ot has a decreased H3K4me3 peak in dysfunctional β-cells when compared to normal β-cells (Figure 58B). Perhaps one model by which *Kcnq1* is being regulated in dysfunctional β-cells is that the presence of a more repressive chromatin environment surrounding *Kcnq1ot* reduces transcription of the *Kcnq1ot* ncRNA, which in turn leads to a reduced repression of the *Kcnq1* gene by creating a more open chromatin state, thereby resulting in an increase in the transcription of the *Kcnq1* gene, ultimately reducing GSIS.

In summary, bivalent marks in normal β-cells are present at genes regulating pancreatic and neuronal development and/or function, and potassium channels that mediate ion permeability. This bivalent state is modified in dysfunctional β-cells and results in gene expression changes that could ultimately contribute to its inability to properly respond to glucose.

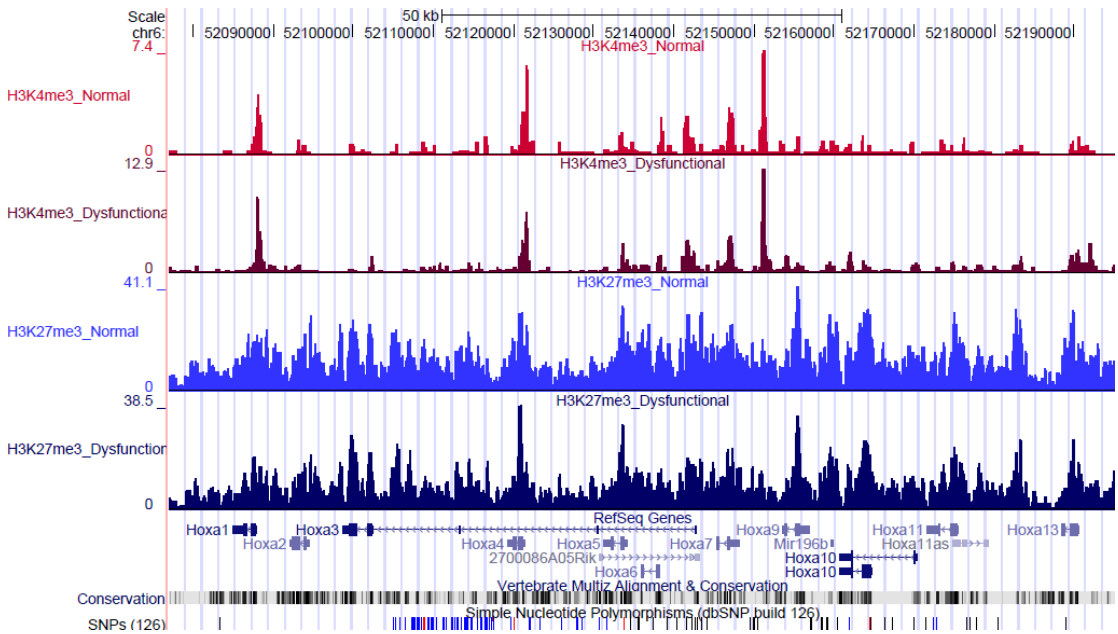
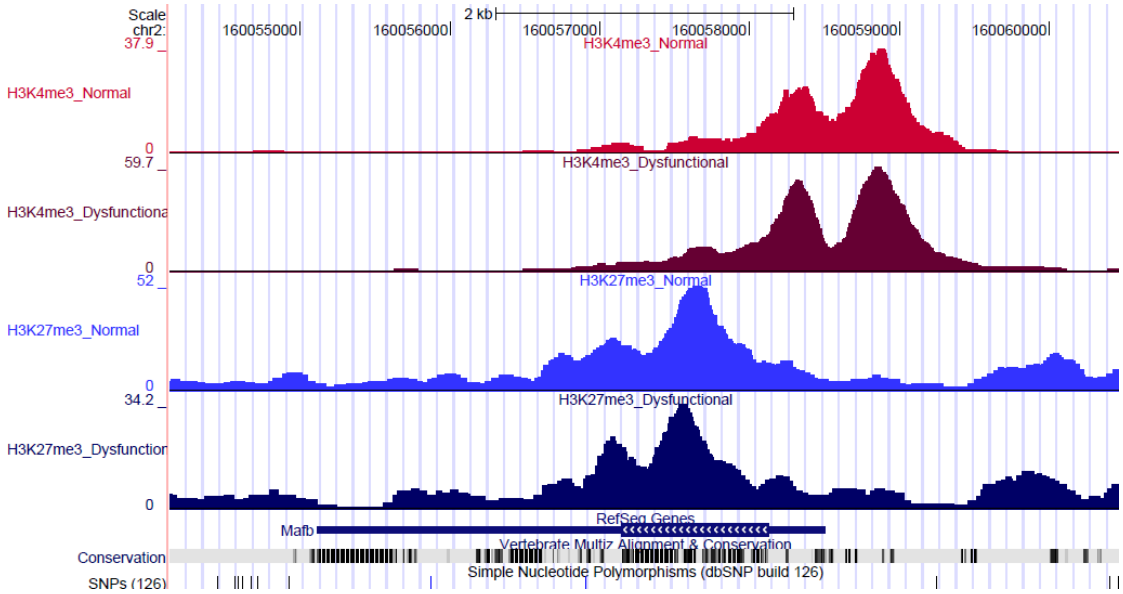
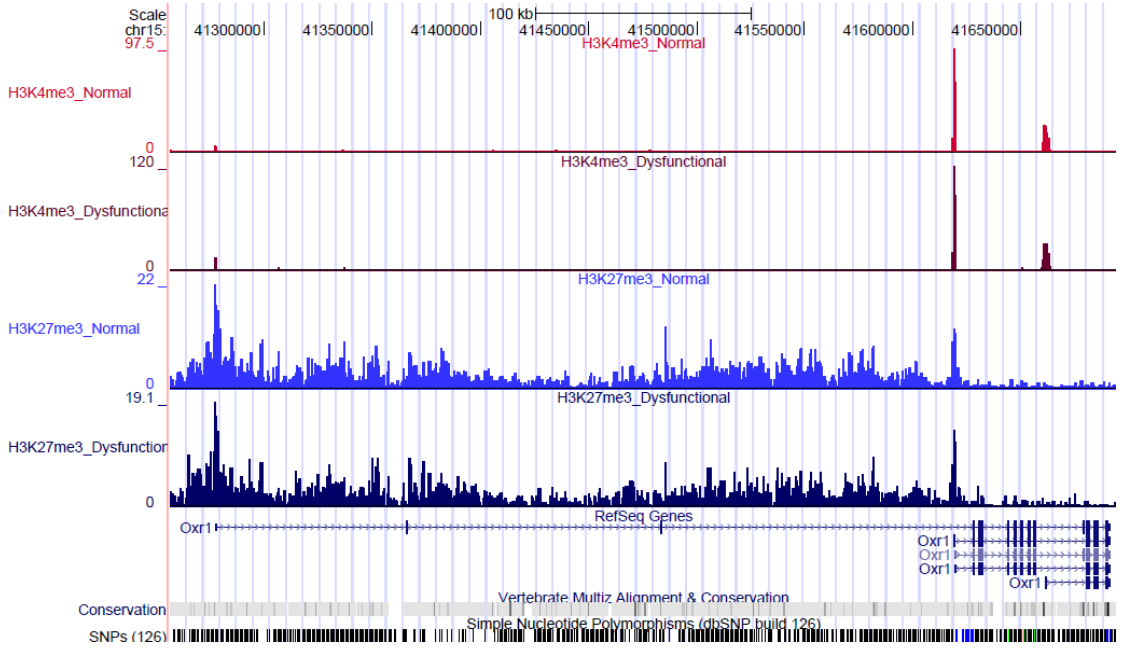


Figure 55. The Hox gene cluster is bivalently marked by H3K4me3 and H3K27me3 levels in normal and dysfunctional β-cells.

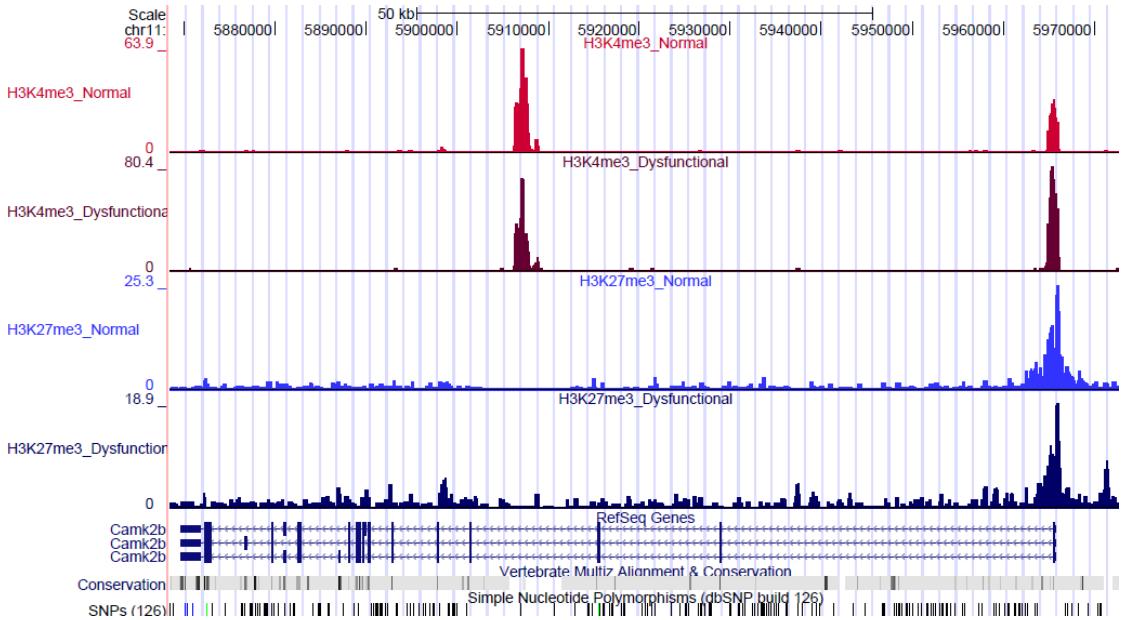
(A) MafB


Scale
chr2:
2 kb
160055000
160056000
160057000
160058000
160059000
160060000
37.9
H3K4me3_Normal
H3K4me3_Normal
0
59.7
H3K4me3_Dysfunctional
H3K4me3_Dysfunctiona
0
52
H3K27me3_Normal
H3K27me3_Normal
0
34.2
H3K27me3_Dysfunctional
H3K27me3_Dysfunctior
0
RefSeq Genes
Mafb
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

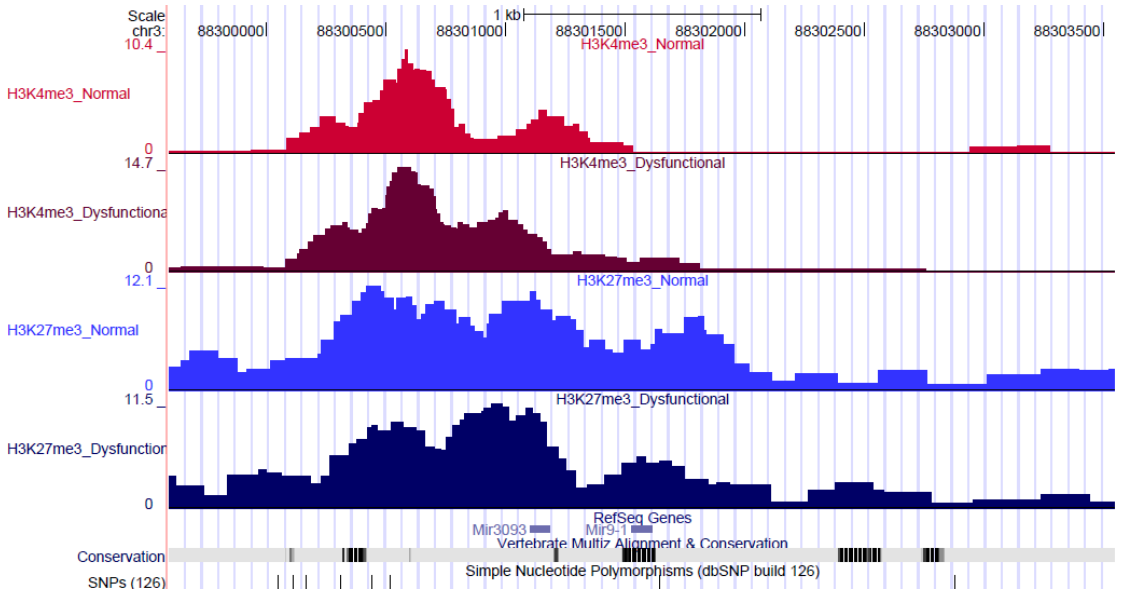
(B) Oxr1


Scale
chr15:
100 kb
41300000
41350000
41400000
41450000
41500000
41550000
41600000
41650000
97.5
H3K4me3_Normal
H3K4me3_Normal
0
120
H3K4me3_Dysfunctional
H3K4me3_Dysfunctiona
0
22
H3K27me3_Normal
H3K27me3_Normal
0
19.1
H3K27me3_Dysfunctional
H3K27me3_Dysfunctior
0
RefSeq Genes
Oxr1
Oxr1
Oxr1
Oxr1
Oxr1
Vertebrate Multiz Alignment & Conservation
Conservation
Simple Nucleotide Polymorphisms (dbSNP build 126)
SNPs (126)

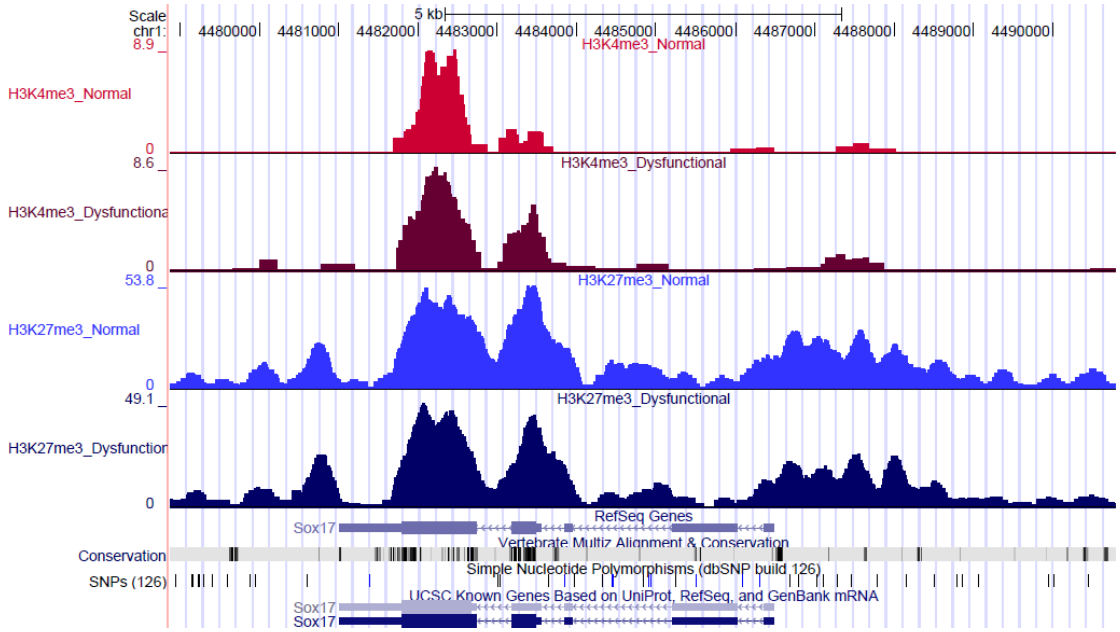
(C) Camk2b



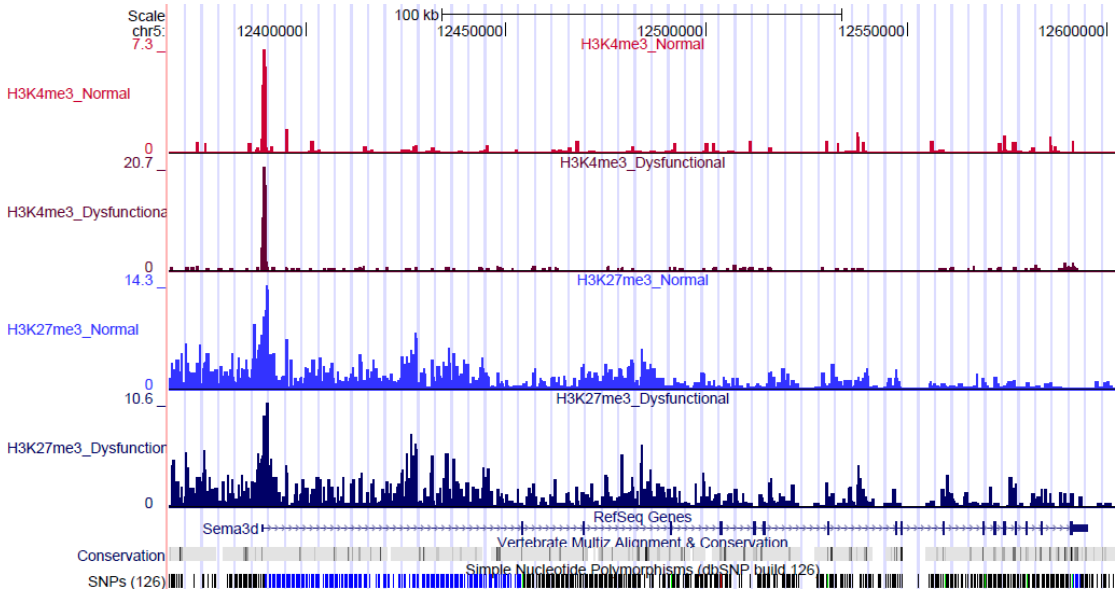
(D) miR-9



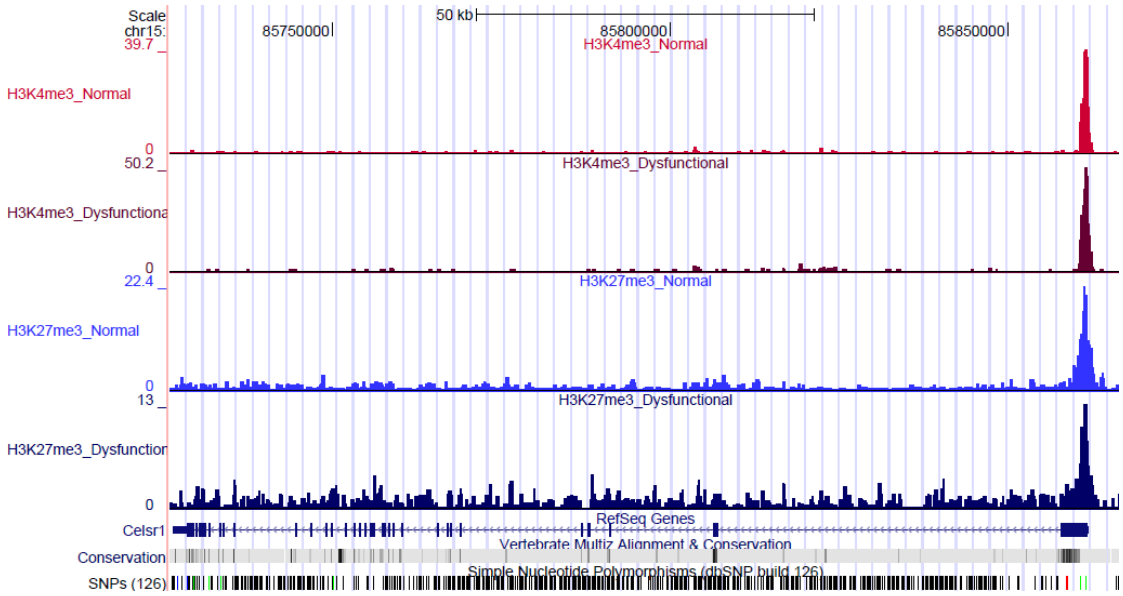
(E) Sox17



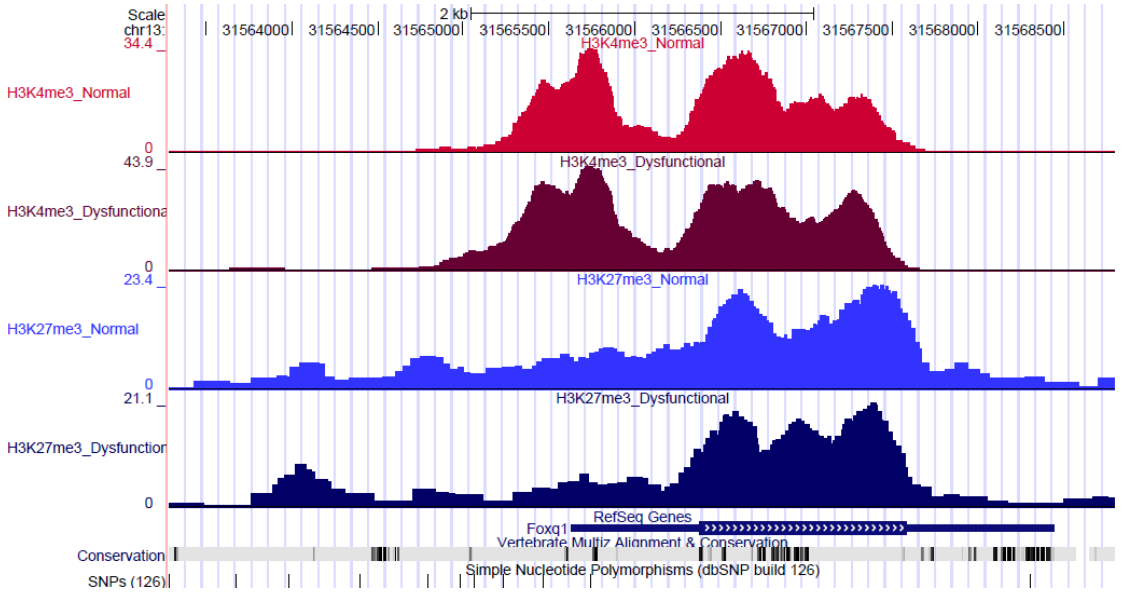
(F) Sema3d



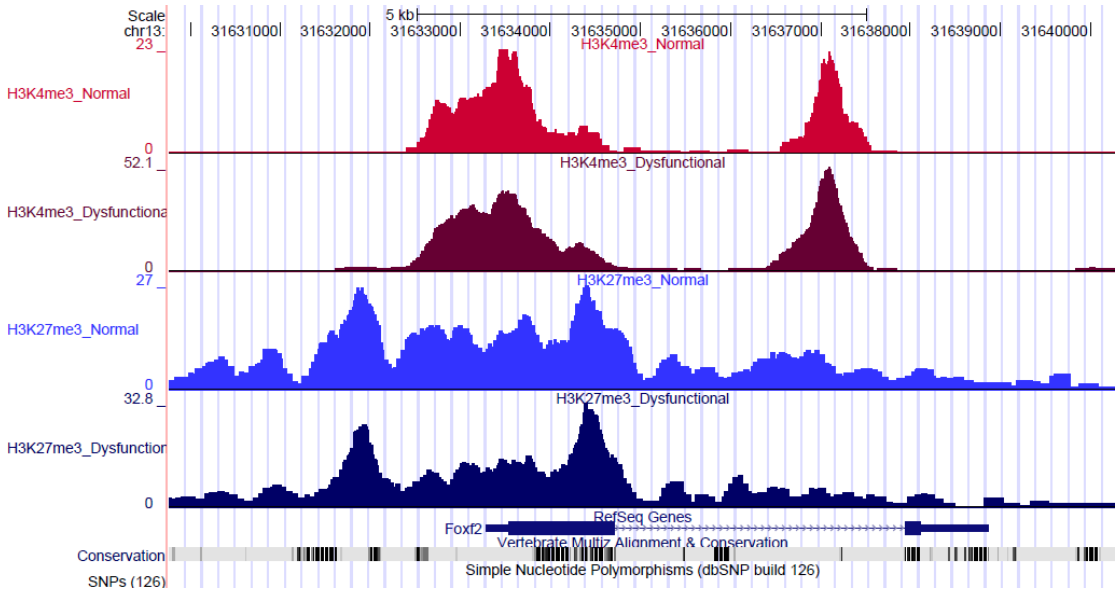
(G) Celsr1



(H) Foxq1



(I) Foxf2



(J) Kcnc1

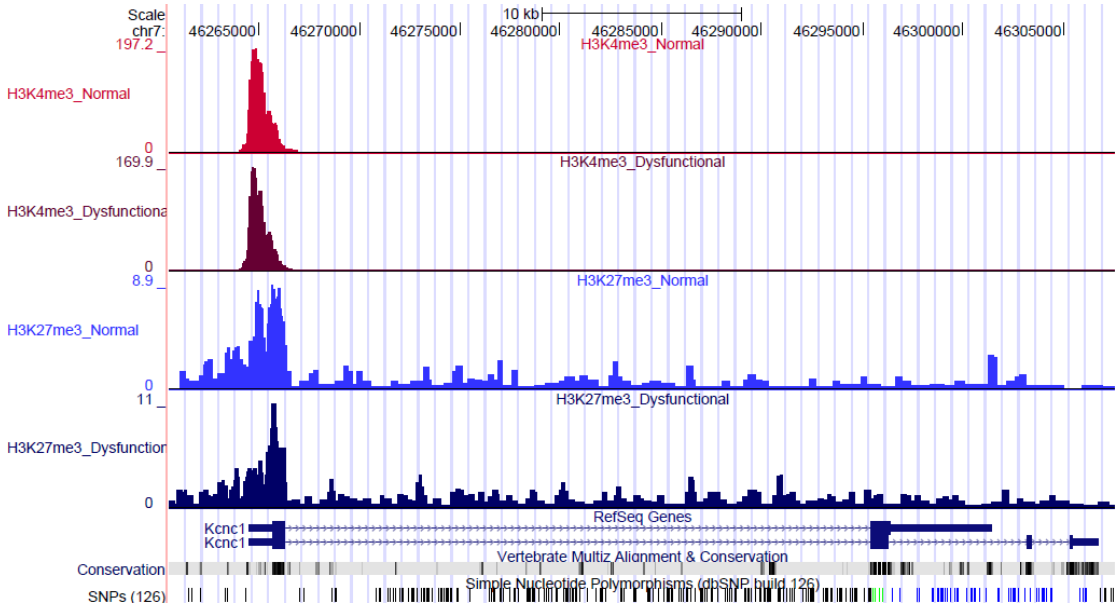


Figure 56. In dysfunctional β-cells, bivalent genes regulating pancreatic development and/or function, neuronal and other developmental lineages, and ion channels, are marked by changes in H3K4me3 and H3K27me3 at the promoter. Bivalent genes in normal β-cells that lost H3K27me3 in dysfunctional β-cells include those involved in pancreatic development and/or function, e.g (A) *MafB*, (B) *Oxr1*, (C) *Camk2b*, (D) *miR-9*, (E) *Sox17*; genes involved in neuronal development and function, e.g. (F) *Sema3d* and (G) *Celsr1*; members of the Fox family transcription factors (H) *Foxf2* and (I) *Foxq1*; and (J) the potassium channel *Kcnc1*.

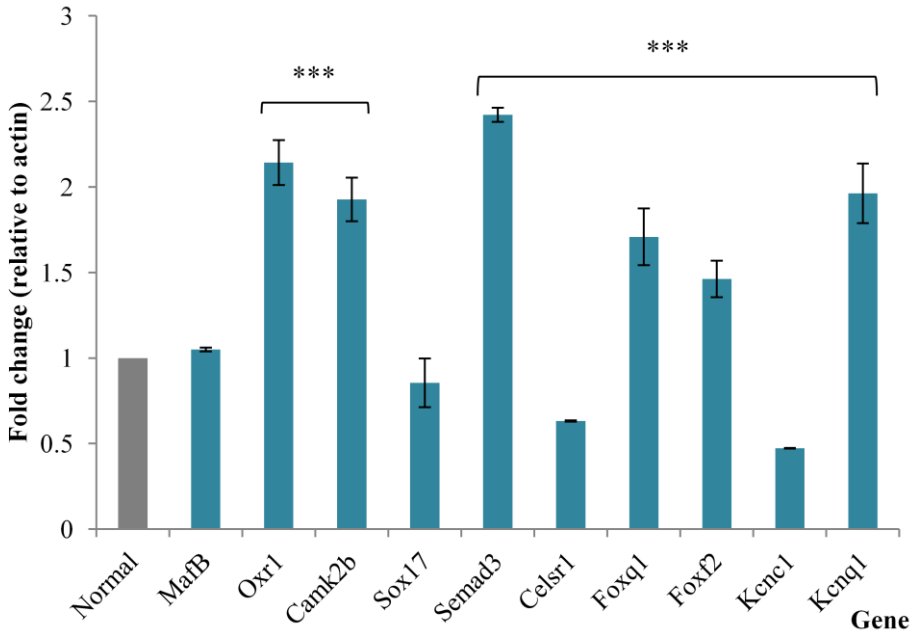
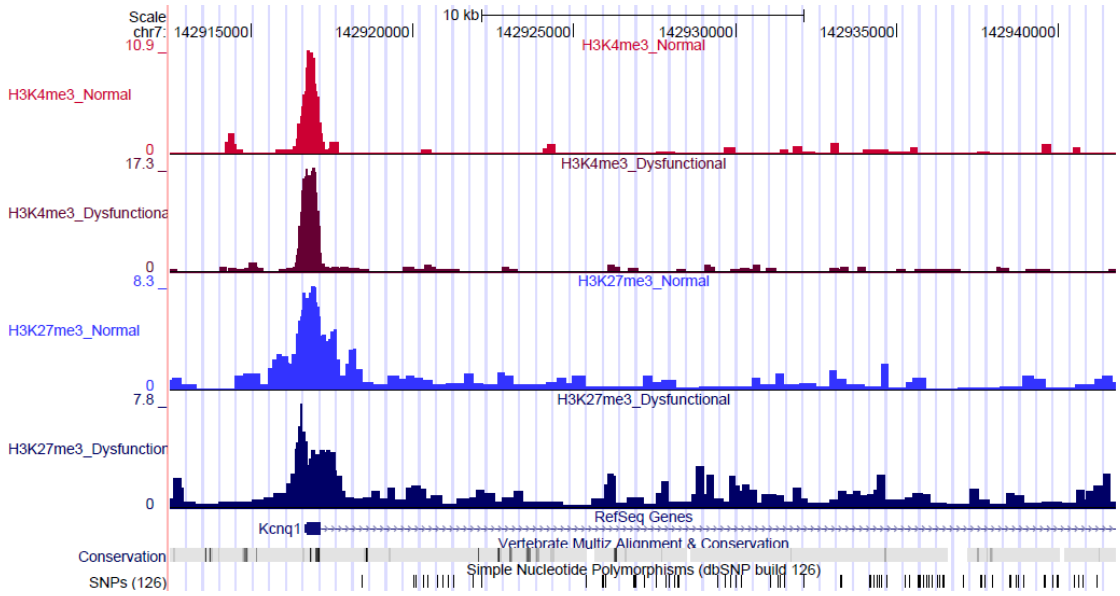


Figure 57. Expression of bivalent genes regulating pancreatic development and/or function, neuronal and other developmental lineages, and ion channels are changing in dysfunctional β-cells. NIT-1 β-cells were treated with 20mM glucose and 0.4mM palmitate for 48hr and RNA harvested from untreated and treated cells (n=4 each). Transcript levels of these genes were assayed by RT-PCR, mean values ± S.E. are plotted as fold change relative to the level of the corresponding gene expression in the untreated control, normalized to endogenous *β-actin*. ***$p \leq 0.02$.

(A) Kcnq1



(B) Kcnq1ot

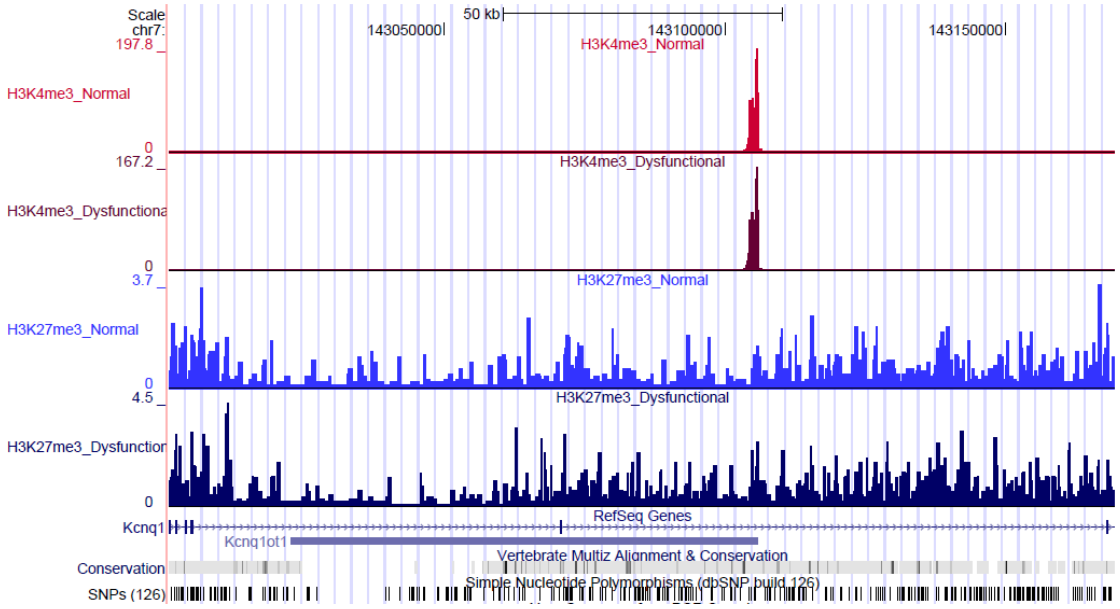


Figure 58. *Kcnq1* (a) and *Kcnq1ot* (b) have opposing bivalent marks in normal β-cells that change in an opposing fashion in dysfunctional β-cells.

5.3. DISCUSSION

5.3.1. CHROMATIN REMODELLING IN RESPONSE TO BETA-CELL DYSFUNCTION

Epigenetic changes are associated with physiological dysfunction due to environmental stimuli. Histone methylation is significantly increased in non-transcribed regulatory regions near or within human Type 1 Diabetes susceptibility regions, including loci containing the gene for CLTA4 and genes in inflammation-associated pathways[312]. The activation of β-cell genes requires an appropriate euchromatin DNA template characterized by hyper-methylation of H3K4, maintained by the HMT Set7/9[162].Therefore, I hypothesized that there would be chromatin remodelling in β-cells in response to a glucolipotoxic stimulus that impairs its ability to secrete insulin in response to glucose.

There was a global increase of chromatin opening in dysfunctional β-cells, as indicated by the increase in numbers of FAIRE and H3K4me3 peaks at the promoters of genes, and a decrease in H3K27me3 at the promoters of genes in dysfunctional β-cells (Table 10). Motif analysis of FAIRE peaks in normal and dysfunctional β-cells showed that there was a bias for binding sites for general transcription factors like SP-1, CTCF, NF-Y (Table 11). SP-1 belongs to a family of ubiquitously expressed, C_2H_2-type zinc finger-containing DNA binding proteins that activate or repress transcription of many genes in response to physiological and pathological stimuli[313]. There is emerging evidence to indicate that in addition to functioning as a 'housekeeping' transcription factor, Sp1 may mediate nuclear signaling in response to hormones such as insulin[314]. CTCF plays many roles in transcriptional repression, insulator function, and imprinting genetic information[315]. NF-Y is a ubiquitous transcription factor involved in the activation of most promoters[316]. Motifs for FAIRE peaks in normal β-cells showed enrichments for tissue-specific transcription factors like FoxA1 and FoxA2, Oct2 and Oct4, and Pdx1 (Table 11). FoxA1 and FoxA2 are important for pancreatic function, with *FoxA1*$^{-/-}$ islets exhibiting decreased GSIS and reduced pancreatic insulin and glucagon content[317]. FoxA2 is important in pancreatic development and function, and regulates many

genes involved in maintaining the mature β-cell phenotype[62]. FoxA1 and FoxA2 co-occupy multiple regulatory domains in the *Pdx1* gene, and ablation of both Foxa1 and Foxa2 in the pancreatic primordium results in complete loss of *Pdx1* expression and severe pancreatic hypoplasia[318]. Oct2 was originally described in cells of the immune system, and plays a regulatory role in the neuro-endocrine brain[319]. Oct4 is a key regulator of pluripotency in embryonic stem cells[320], and high expression of *Oct4* and *Nanog homeobox (Nanog)* in metaplastic pancreatic ducts (with *Oct4* expression preceding Ras mutation) suggests that these homeobox transcription factors are associated with the early stage of pancreatic cancer carcinogenesis and may play an important role in that process[321]. The Usf1 motif was one of the top 20 motifs identified in the normal β-cell FAIRE datasets. Usf1 and NeuroD1 contribute to *islet-specific glucose-6-phosphatase catalytic-subunit-related protein (IRGP)* gene expression in islets[110]. Taken together, this data shows that in dysfunctional β-cells, open chromatin is biased to binding sites for general transcription factors; in normal cells, tissue-specific TFBS tend to show a more open chromatin configuration. Comparison of the genes associated with the respective epigenetic marks in normal and dysfunctional β-cells demonstrated that there is a significant increase in genes involved in cellular regulatory activities and the specification of other developmental lineages in dysfunctional β-cells (Figures 44-6).

To summarize, I have shown that important genes in normal β-cells are marked by active chromatin marks, and there is global chromatin remodelling when β-cells become dysfunctional, which results in an opening in the chromatin at genes involved in cellular regulatory activities and the specification of other developmental lineages.

5.3.2. CHROMATIN AND TRANSCRIPTIONAL DYNAMICS IN DYSFUNCTIONAL BETA-CELLS

Genes important for β-cell function are marked by H3K4me3 in their promoters in normal β-cells, e.g. *Pdx1, NeuroD1, MafA, Ins1, Glut2, Snap25, Vamp2, Cacna1a, Chgb, Kirrel6.2, Sytl4, Stxbp1,* and *Syt7* (Figure 47). This is in stark contrast to two prior studies of H3K4me3 ChIP-Seq in human pancreatic islets, where the authors did not find the H3K4me3 mark in highly expressed genes in islets, including those encoding islet-specific hormones Ins, *Gcg, Sst, Iapp, Ppy* and *Ttr*[140,179]. This prompted the authors to speculate that some genes critical for islet function have an unconventional promoter chromatin signature, indicative of a unique transcriptional control mechanism[176]. However, the fact that there were positive enrichments for H3K4me3 at so many important β-cell genes in my study does not support their conclusion. Perhaps this is due to species differences - they used pancreatic human islets while I used mouse insulinoma cell lines. The fact they used whole pancreatic islets from cadaveric donors (islets are extremely fragile and sensitive to external manipulation, and islet harvesting is a laborious process that triggers a cascade of stressful events involving activation of apoptosis and necrosis and the production of pro-inflammatory molecules that negatively influence islet yield and function)[322] might have also compounded the results. β-cell paucity of their samples might have also contributed to their conflicting results, and there is heterogeneity when using whole islets that contain α, β, δ, ε, and PP cells.

Under glucolipotoxic conditions in DIO models[168] and cell culture conditions mimicking diabetes[169], *MafA* expression is inhibited, whereas Pdx1 is affected at the post posttranslational level in its ability to translocate to the nucleus[280], resulting in decreased *Insulin* gene expression. It is not known what effect glucolipotoxic conditions have on NeuroD1, but I have shown there is a decrease in mRNA levels of all three transcription factors in dysfunctional β-cells (Figure 49). Glucotoxicity and lipotoxicity leads to the production of ROS, which activate *Mitogen-activated protein kinase 8 (Jnk),* resulting in a decrease in Insulin receptor substrate (Irs) signalling, which may directly be involved in

decreased Pdx1 activity by translocation from the nucleus to the cytoplasm[323]. Glucose and FFA also induce ER stress, while chronic glucose elevation inhibits FFA oxidation and favors the generation of ceramide and lipid partitioning, which also results in β-cell dysfunction[324].

Genes bound by Pdx1, NeuroD1 and MafA in normal β-cells are associated with H3K4me3, not H3K27me3 (Figure 50). There is an increase in the association of these transcription factors with H3K27me3 in dysfunctional β-cells and a corresponding decrease in gene expression (Figure 52). Taking into account there was reduced expression of *Pdx1*, *NeuroD1* and *MafA* in dysfunctional β-cells (Figure 49), the H3K27me3-marked genes (that have a TFBS) in dysfunctional β-cells are enriched for genes involved in metabolic and general cellular regulatory processes (e.g. cell cycle, cell signalling) (Tables 12-14). In the case of Pdx1, where there also is (in theory) reduced Pdx1 translocation to the nucleus, the increase in H3K27me3 of its target genes could also be due to the presence of less Pdx1 in the nucleus.

Taken together, two things are happening: Transcriptional activity of *Pdx1*, *NeuroD1* and *MafA* are reduced, and so are those of their target genes. The former is due to the high glucose and fatty acid conditions leading to epigenetic changes in their promoters (as evidenced by reduced H3K4me3 levels at the *Pdx1*, *NeuroD1* and *MafA* promoters in dysfunctional β-cells, Figure 47). The latter has two possible explanations: As a consequence of treatment with high glucose and fatty acids, there is a decrease in the level of the active H3K4me3 mark at the promoters of *Pdx1*, *NeuroD1* and *MafA*, leading to their decrease in gene activity. This then directly leads to epigenetic changes in the promoters of their target genes, resulting in changes in gene expression. The second possibility is that conditions of high glucose and fatty acids lead to epigenetic changes at the promoters of *Pdx1*, *NeuroD1* and *MafA* target genes, on top of epigenetic changes at the promoters of those three transcription factors, resulting in changes in gene expression. The experiments I have performed cannot distinguish between these two possible scenarios. However, since chomratin modifiers are usually not sequence-specific, and that Pdx1, NeuroD1 and MafA interact with HATs, HMTs and HDACs in order to maintain

the expression of *Insulin* and other genes, it is unlikely that chromatin modifiers would target specific gene sets. Hence, the first scenario is more likely.

I therefore propose a model in which glucolipotoxic conditions lead to increase in ROS and other metabolic changes that result in a reduced H3K4me3 mark at the *Pdx1*, *NeuroD1* and *MafA* genes and correspondingly lower mRNA levels in dysfunctional β-cells. This in turn leads to chromatin remodelling at the promoters of their target genes, resulting in changes in gene expression that ultimately affect β-cell function. This is similar to the model that is proposed to occur in IUGR rodents, where a cascade of epigenetic events is triggered by IUGR during development, resulting in permanent suppression of *Pdx1* expression[167].

ChIPs for these transcription factors can be performed in dysfunctional β-cells, and if reduced levels of binding are seen when compared to normal cells, it would lend support to the model I have proposed. There are several other important transcription factors involved in maintaining the mature β-cell phenotype, e.g. FoxA2, Hnf4α; these also interact with chromatin modifiers to modulate gene expression[325]. Therefore, ChIPs for those transcription factors could be performed in normal and dysfunctional β-cells, to see if a similar form of regulation is occurring. ChIPs of other epigenetic marks and/or HATs/ HMTs/ HDACs can also be performed under varying conditions of glucolipotoxicity, in order to tease apart the mechanistic effects underlying β-cell dysfunction. DNA methylation at the promoters of important β-cell genes can analyzed to shed light on how DNA methylation and histone modification co-operate to modulate changes in gene expression. These experiments can also be performed in DIO animals or human pancreatic islets from normal and diabetic patients, to see if similar changes are observed *in vivo*.

5.3.3. Bivalent States in Normal and Dysfunctional Beta-Cells

Bivalent opposing marks are associated with low levels of expression in pluripotent cells, but these resolve into monovalent methylation marks during ESC differentiation. Bivalent genes

in ESCs are mostly transcription factors, hence the notion that bivalent marks silence genes in ESCs, priming them for activation upon differentiation[137]. Bivalent domains are also found in other cell types, e.g. mouse embryonic fibroblasts, neural progenitors[138] and human T cells[139]. A recent comparison of H3K4me3 and H3K27me3 in human pancreatic islets show that these bivalent domains are usually mutually exclusive or absent, but there are a few genes that had high levels of both marks, e.g. the *Hox* gene clusters and neuronal transcription factors, perhaps indicating a retention of "memory" of earlier transcription or the re-activation of these genes under certain metabolic conditions[140]. Bivalent marks have also been described to co-exist with methylated DNA in the promoter regions of human cancer cells, suggesting that aberrant *de novo* DNA methylation that commonly accompanies cancer-related genes could be a consequence of the underlying chromatin environment[326].

When H3K4me3 and H3K27me3 overlapping peaks in normal and dysfunctional β-cells were analyzed, there were 113 bivalent genes in normal β-cells. There were 16 genes with lower H3K4me3 mark in dysfunctional β-cells and 98 genes with a lower H3K27me3 mark in dysfunctional β-cells. This is consistent with the fact that there was an increase in the numbers of H3K4me3 and FAIRE peaks, and decrease in the global numbers of H3K27me3 peaks in dysfunctional β-cells. Genes that are bivalently marked in normal β-cells are more enriched for in developmental processes and cell regulatory processes, compared to genes that are bivalently marked in normal β-cells and have lost the repressive H3K27me3 mark in dysfunctional β-cells (Figure 54).

Some examples of bivalent genes in normal β-cells that have reduced levels of H3K4me3 and/or increased levels of H3K27me3 in dysfunctional β-cells include genes involved in pancreatic development and/or function, e.g. *MafB, Oxr1, Camk2b, miR-9, Sox17* (Figure 56). Genes involved in neuronal development and function, e.g. *Sema3d* and *Celsr1*, and members of the *Fox* family transcription factors *Foxf2* and *Foxq1* are also bivalently marked (Figure 56). The potassium channel *Kcnc1* has both active and repressive chromatin marks, along with

the T2DM-identified potassium channel *Kcnq1* (Figure 58). Bivalent chromatin states are modified to different extents in dysfunctional β-cells and lead to gene expression changes that could ultimately contribute to its inability to properly respond to glucose.

5.4. SUMMARY

My hypothesis was that when β-cell dysfunction occurs, remodelling of the chromatin environment adversely affects the expression and activity of Pdx1, MafA and NeuroD1 and other important β-cell transcriptional regulators. I have addressed this by looking at regions of active and repressive chromatin in normal and dysfunctional β-cells, and found that dysfunctional β-cells have a more open chromatin configuration marking genes involved in cellular regulatory activities and the specification of other developmental lineages. I have also found an increase in the number of transcription factor-associated genes that bear an increased H3K27me3 mark in dysfunctional β-cells. Bivalent genes are present in normal β-cells, and a change in the bivalent state is observed at genes involved in development and/or function. This chromatin remodelling leads to changes in gene expression that could ultimately contribute to its inability to properly respond to glucose.

CHAPTER 6:

OVER-EXPRESSION OF PDX1 DURING PANCREATIC BETA-CELL DIFFERENTIATION

6.1. INTRODUCTION

ESCs are derived from the pluripotent inncer cell mass of early embryos and possess a nearly unlimited capacity for self-renewal and the potential to differentiate into the many cell types of mammalian organisms. These properties qualify ESCs as an excellent system to analyze *in vitro* lineage commitment and cellular differentiation. Though mouse ESCs (mESCs) and human ESCs (hESCs) are not developmentally identical, clues obtained from experimental work performed in mESCs might shed light towards the application of mouse differentiation systems to hESCs, and might therefore lead to the establishment of strategies for future regenerative cell therapies. One strategy to generate β-cells *in vitro* is to closely mimic embryonic development by exposing ESCs to factors (e.g. transcription factors, growth factors, signaling molecules, the appropriate extra-cellular matrix (ECM) environment) that they would normally encounter *in vivo*[327].

In vitro recapitulation of β-cell development has been attempted by: (1) directed differentiation using lineage selection with genetically modified ES cells that express pancreas-specific promoter and reporter system[328,329], (2) application of specific growth and ECM factors during differentiation[330,331] and (3) constitutive expression of genes crucial for pancreatic development[328,332]. However, currently strategies require optimization with respect to maturation of insulin-producing cells and the ability of these differentiated cells to respond to glucose when transplanted in diabetic animal models. Therefore, promising future strategies include the conditional expression of transgenes, histotypic differentiation supported by vascular endothelial cells and ECM-based three-dimensional culture systems, and application of physiological-like differentiation conditions.

Pdx1 has been shown to be essential in the development of the pancreas, with deficits in expression levels resulting in reduced insulin secretion, accelerated β-cell apoptosis, and insulin-deficient diabetes in mice[81]. I have also shown in Sections 3-5 that Pdx1 is

comparatively more important than other key regulators like NeuroD1 and MafA. Thus, I hypothesized that **over-expression of Pdx1, in combination with culture conditions that mimic endoderm and pancreas development, will aid in the differentiation of mouse ESCs towards the pancreatic lineage.**

6.2. RESULTS

6.2.1. GENERATION OF MOUSE CLONAL ESC LINES THAT OVER-EXPRESS PDX1

An inducible mouse ESC line over-expressing *Pdx1* was generated using the E14 engineered cell line AINV18[333]. The objective was to expose it to several differentiation conditions and to examine gene expression levels as a first step in determining efficiency of differentiation. In this system, the parental AINV18 ESC line contains an expression cassette for the reverse tet transactivator (rtTA) inserted into the constitutively expressed ROSA26 locus (Figure 59A). The cDNA of interest, in this case *Pdx1,* is targeted by Cre-mediated homologous recombination downstream of the tetracycline (tet) operator. This results in Pdx1 being under the control of the tet operator, which can be switched on upon addition of doxycycline (a tet analog). This recombination event also places a previously inactive neomycin resistance gene under the control of a PGK-ATG promoter, allowing selection of Pdx1 recombination with neomycin.

With this method, I generated two ESC clones over-expressing Pdx1 that were able to switch on *Pdx1* expression at the mRNA and protein level within 3 hours, maintain robust expression over a period of 12 days, and switch off *Pdx1* expression upon withdrawal of doxycycline (Figure 59B).

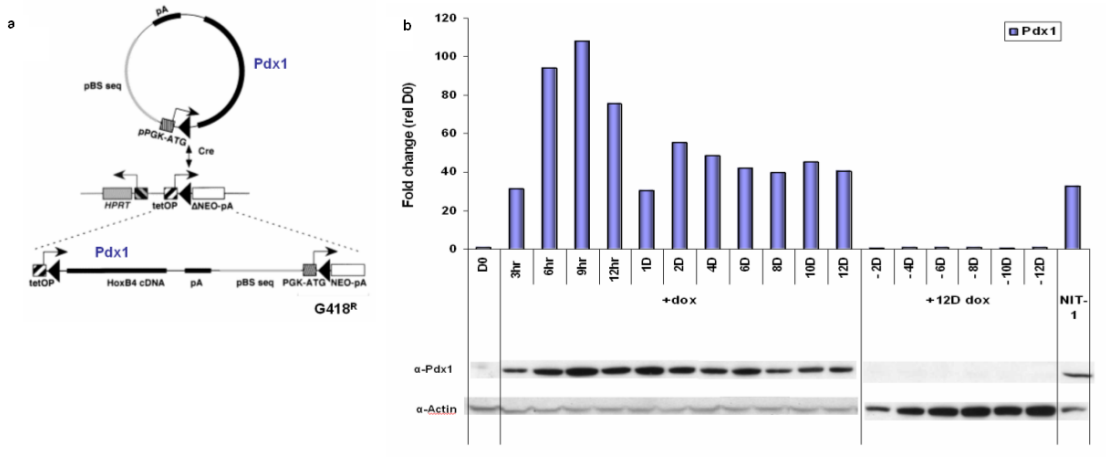


Figure 59. Co-transfection of pLox containing Pdx1 with Cre causes Cre-mediated recombination of Pdx1, now under the control of the tetOP, and G418 resistance is established. (B) Quantitative RT-PCR analysis of *Pdx1* mRNA expression (upper panel) and protein levels (lower panel) before and after addition of doxycycline, with NIT-1, a β-cell line serving as a positive control. Similar results were obtained with the second Pdx1 ESC clone.

6.2.2. DIRECTED DIFFERENTIATION OF MESCS TO THE BETA-CELL LINEAGE

Using a differentiation protocol that was used to over-express *Ngn3* in a similarly inducible manner[334], I differentiated the parental AINV18 cell line and looked at gene expression changes of several key genes involved in maintaining pluripotency, formation of definitive endoderm, pre-pancreatic endoderm, endocrine precursors and β-cells (Figure 60A-F).

In stage 1, ESC differentiation was initiated by generating embryoid bodies, to allow cells to form a three-dimensional structure that allows intra-cellular signaling. Leukemia inhibitory factor (LIF) was withdrawn to promote differentiation, as seen by the decrease in the pluripotency marker *Oct4* and *Sox2* (Figure 60A). In Stage 2, serum was withdrawn and Nodal signaling by adding Activin A and Fibroblast growth factor (FGF)-4 to potentiate endoderm formation and anterior gut fate. This resulted in up-regulation of *FoxA2*, *Sox17*, *Gata4* and *Gata6* (Figure 60B), markers of primitive endoderm. In stage 3, Activin A and FGF4 were maintained in the culture medium and bone morphogenic protein (BMP)-4 and FGF10 were added to expand the population of putative pancreatic progenitors. This effect was confirmed by maintenance of *FoxA2*, *Sox17*, *Gata4* and *Gata6* levels (Figure 60B). Pancreas specification depends on addition of RA and suppression of *Shh* by cyclopamine, which represses the mesodermal markers *Gsc* and induces expression of *Pdx1* (Figure 60C), *Ngn3* (Figure 60D) and *Ptf1a*, as detected by Stage 5. Low glucose was used in stages 4 and 5 because glucose restriction may initiate generation of insulin producing cells[335]. Nicotinamide, a SIRT-specific inhibitor, is commonly used as part of the differentiation cocktail to drive ESCs towards the β-cell lineage[330]. This protocol was modified with the addition of β-cellulin, a glycoprotein highly expressed in the pancreas, to promote the differentiation, regeneration and proliferation of pancreatic β-cells *in vitro* conditions[186]. Addition of β-cellulin also enhances the expression of *Pdx1*, *Ngn3*, *Nkx6.1*, *Insulin* (Figure 60C-F) and other genes specifying pancreatic progenitors. Therefore, with this protocol, a population of cells expressing markers representing β-cells e.g. *Nkx6.1* (Fiugure 60E) and *Insulin* (Figure 60F) was produced.

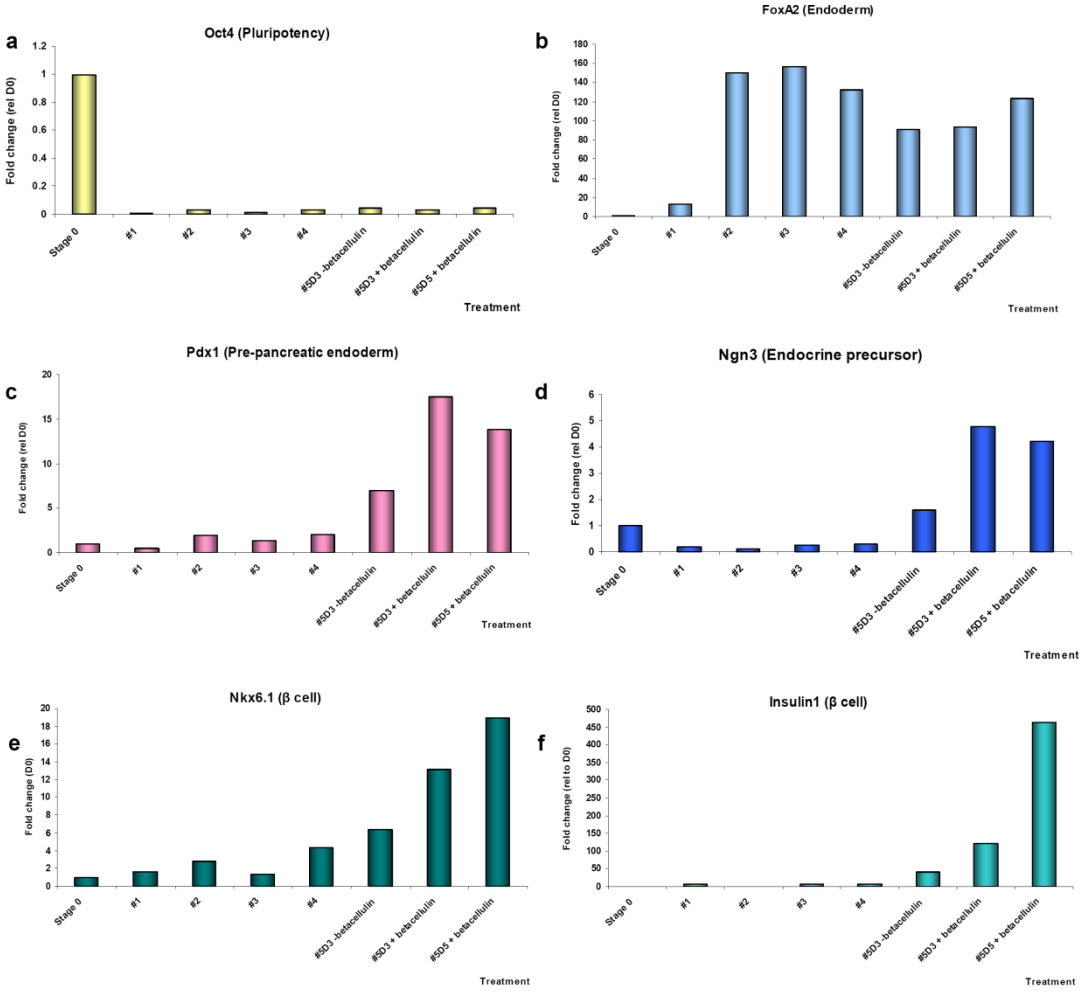


Figure 60. Expression of marker genes at each stage of ESC differentiation towards the β-cell lineage. β-actin was used as an internal control for normalization. Three technical replicates were performed for each RT-PCR, and the average of the normalized ratio of gene/β-actin was calculated. Similar results were obtained with the second Pdx1 ESC clone.

2.3. Directed Differentiation of mESCs Towards the Beta-cell Lineage, in Combination With Pdx1 Over-Expression

As *Pdx1* expression was detected at stage 5 of the differentiation protocol (Figure 60C), I asked whether over-expressing *Pdx1* before, during, or after stage 5 would enhance the differentiation process. I induced *Pdx1* expression by addition of doxycycline to the differentiation medium at stages 4, at stage 5, and midway through stage 5 (Figure 61A). Marker gene expression was assessed at each stage of differentiation by RT-PCR (Figure 61B-F). The experiment was performed with both Pdx1 ESC clones, and results shown in Figure 63 are representative of one clone, with the trend in the expression results also seen in the other clone. I assayed a panel of markers for each differentiation stage, i.e., pluripotency, formation of definitive endoderm, pre-pancreatic endoderm, endocrine precursors and β-cells, at each stage in the protocol. The controls for the experiment are the parental AINV18 cell line without any genetic modification and the Pdx1 clone without doxycycline treatment.

Levels of *Oct4* transcripts decreased below that of control cells at the end of the differentiation experiment, with some remnants of *Oct4* present in the *Pdx1* over-expressing clone (Figure 61B), indicative of cells still remaining undifferentiated. *Pdx1* transcription was massively up-regulated upon addition of doxycycline (Figure 61C), showing that the cells were responsive to induction of doxycycline in the context of the differentiation medium. Levels of *Ngn3* transcripts were not increased upon *Pdx1* over-expression (Figure 61D), nor was *Nkx6.1* (Figure 61E) and *Insulin* (Figure 61F). In fact, the levels of *Ngn3*, *Nkx6.1* and *Insulin* (and other markers representing pancreatic progenitors, data not shown) were higher in the AINV18 parental cell line and in the Pdx1 clonal ESCs without the addition of doxycycline. The data suggests that over-expression of *Pdx1* did not aid differentiation towards the β-cell lineage, but seems to impede it. This almost seems counter-intuitive, as one would expect, with Pdx1 being upstream of these transcription factors and being crucial for β-cell development, over-expressing *Pdx1* would lead to an increase in the levels of pancreatic progenitor transcripts, not decrease.

(A)

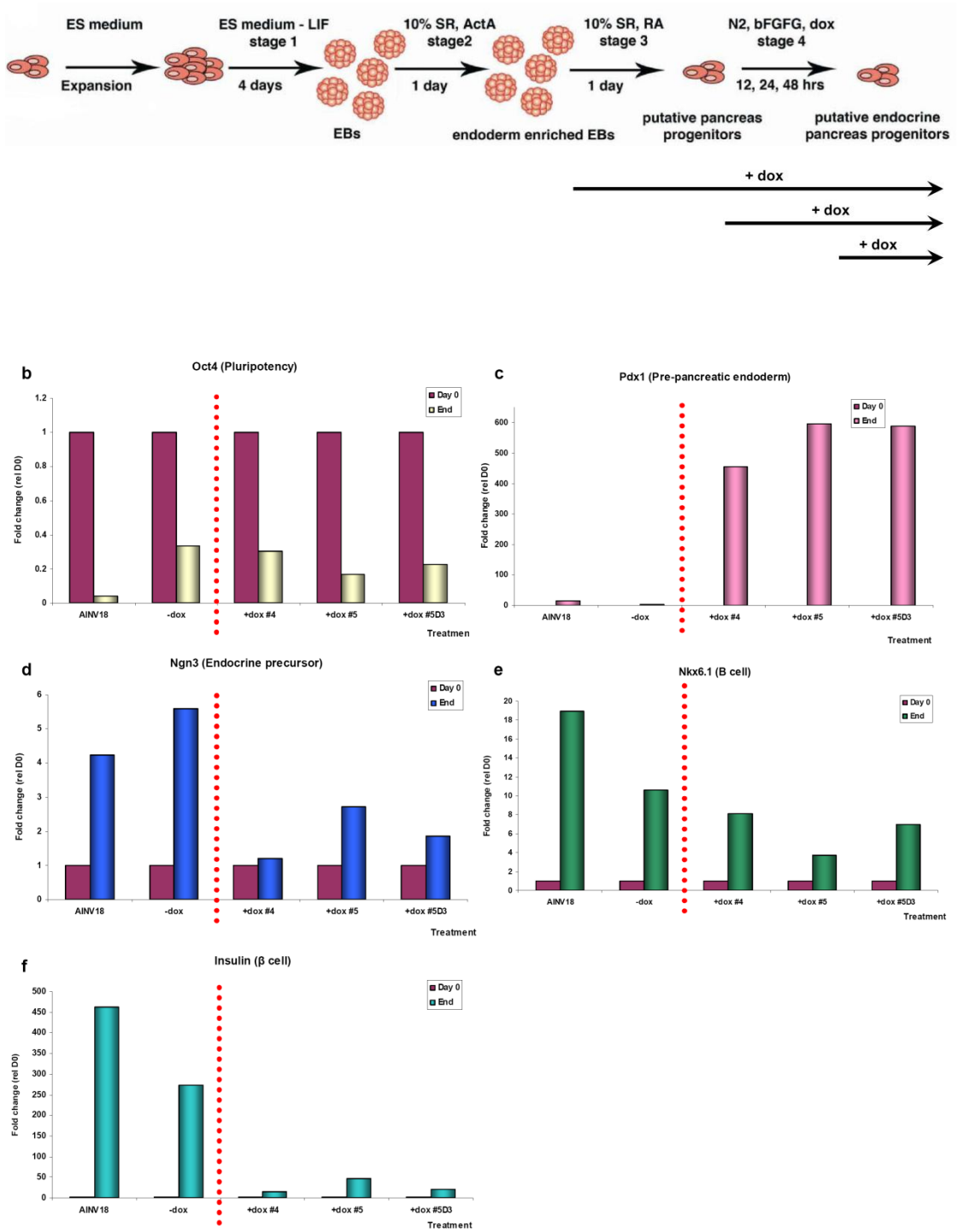


Figure 61. Pdx1 over-expression does not enhance differentiation of ESCs towards the endocrine lineage. (A) Differentiation of ESCs towards endocrine pancreas-like progenitors. A schematic diagram of the differentiation protocol and Pdx1 induction. (B) to (E) shows expression of markers from each stage of differentiation. Data shown are analyses performed at the end of stage 5 on embryoid bodies that had doxycycline added at stage 4, stage 5, or midway through stage 5. The parental AINV18 cell line and the Pdx1 cell line (without dox) served as controls, and for each RT-PCR three replicates were averaged. Results were similar for both Pdx1 ESC clones.

6.3. DISCUSSION

Embryonic stem cells hold great promise for therapy as they can differentiate into any cell type in the body. As such, much effort has been invested in enhancing the derivation of different cell types, such as the pancreatic β-cell. I have shown that over-expressing Pdx1 does not enhance the differentiation towards the β-cell lineage (Figure 61). This was somewhat surprising, as Pdx1 lies upstream of transcription factors involved in specifying the endoderm and endocrine lineage, my hypothesis was that over-expressing *Pdx1* would lead to an increase in the levels of pancreatic progenitor transcripts, not decrease (Figure 61). It is possible that there was over-saturation of translational machinery when *Pdx1* was over-expressed, as there was a massive increase in the levels of *Pdx1* mRNA transcripts (500-fold) upon addition of doxycycline and, therefore, other mRNA transcripts could not be translated efficiently (Figure 61). This huge increase in *Pdx1* levels could have also led to non-specific and non-physiological binding and spurious activation of its target genes. One strategy to overcome non-physiological levels of *Pdx1* is to switch on *Pdx1* in biphasic pulses (e.g. for several hours in Stage 4, and midway through Stage 5 for a couple of hours, to mimic the biphasic expression levels of Pdx1 levels in the developing embryo).

Immuno-staining for C-peptide, Insulin, and the other β-cell markers should also be performed on differentiated cells, to determine the protein levels and co-localization of factors characteristic of β-cells, as I've only looked at mRNA transcripts so far. Fluorescence-activated cell sorting (FACS) will be carried out as a quantitative assay to assess the percentage of cells expressing the relevant markers, for example the endodermal markers Sox17 or FoxA2, to evaluate the efficiency of the differentiation process. Enrichment of differentiated cells expressing endodermal surface markers can also be performed, (e.g. Cxcr4 distinguishes the early visceral and definitive endoderm[336]).

6.3.1. MODIFICATION OF THE ECM

Mature β-cells do not produce any basement membrane or ECM, using instead VEGF-A to attract endothelial cells which form capillaries with a vascular basement membrane next to the β-cells[337]. Laminins and the β1 integrin subunit are crucial for the stimulatory effect of endothelial ECM proteins on *insulin* mRNA expression. Therefore, β-cells must interact with endothelial cells in order to function properly. In dissociated E13.5 precursor pancreatic cell cultures, the addition of soluble laminin-1 increased the number of β-cells produced[338]. Also, β-cells grown with collagen IV or laminin secrete more insulin in response to glucose stimulation[339]. Matrigel[340], an ECM commonly used in culturing human ESCs, contains growth factors like bFGF, EGF, IGF-1, PDGF, NGF, laminin, collagen IV. It is known that long term culture of fetal porcine islet-like cell clusters with Matrigel enhances insulin secretion[341]. As the differentiation protocol can be easily modified, and because the ECM is crucial in providing a proper ECM context, efforts could be made to focus on plating embryoid bodies onto the various ECMs mentioned above (after stage 1 of the differentiation protocol, or midway in the differentiation protocol when cells are more mature) onto various ECMs (laminin 1, collagen IV, matrigel, 804G matrix, and specialized matrices which contain different combinations of ECM proteins). A similar approach would be to co-culture differentiating cells onto mouse islet endothelial cells (e.g. MS-1 cells), which promote β-cell proliferation through the secretion of HGF, suggesting the existence of an endothelial-endocrine axis within adult pancreatic islets important for adult β-cell proliferation[342].

6.3.2. MODIFICATION OF EXOGENOUS COMPONENTS OF DIFFERENTIATION MEDIUM

In order to modify exogenous components of the differentiation cocktail, other growth factors and signaling molecules shown to be important in β-cell development can be added, such as FGF7 and Noggin[27]. Exendin-4 (a Glp-1 analog that increases GSIS from β-cells)[343], can also be supplemented to the culture medium. Under appropriate conditions, the HDAC inhibitor

Trichostatin A (TSA) can differentiate bone marrow stem cells into islet-like clusters capable of secreting insulin[344]. TSA and sodium butyrate (both class I and II HDAC inhibitors) treatment increased the development of Ngn3-positive endocrine progenitors and modified the endocrine sub-type lineage choices[345]. Therefore, there is much to be optimized with respect to differentiation culture conditions.

6.4. SUMMARY

The desire to generate clinically relevant populations of cells from ESCs remains the driving force of ESC research. I have shown that over-expression of Pdx1 alone does not improve differentiation towards the pancreatic β-cell lineage. Future work includes a combinatorial approach, consisting of timely over-expression of a key pancreatic transcription factor with defined culture conditions, where differentiating cells will be exposed to the proper ECM context.

CHAPTER 7:

DISCUSSION

7.1. TRANSCRIPTIONAL REGULATION OF PDX1, NEUROD1 AND MAFA IN PANCREATIC BETA-CELLS

Pancreatic β-cells are important in maintaining glucose homeostasis in the adult, secreting insulin in a timely and appropriate fashion in response to a metabolic stimulus. The inability of the β-cell to keep up with increasing insulin demand, coupled with insulin resistance in other parts of the body, leads to Type 2 Diabetes. It is therefore of scientific interest to understand how these highly specialized cells perform their function, and to also understand the molecular mechanisms in which β-cell dysfunction occurs. The transcriptional and epigenetic networks governing the unique properties of β-cells has not been extensively studied as yet, with only a handful of publications describing genome-wide profiling of transcription factors and the epigenetic profile of β-cells[86,177,179,325]. The transcription factors Pdx1, NeuroD1 and MafA are required for the maintenance of the mature β-cell phenotype, with deficiencies in any of them leading to diabetic phenotypes[1,195,346]. I have thus first focussed on further elucidating the transcriptional networks regulated by them.

Pdx1 and NeuroD1 co-occupy hundreds of genomic loci[86], implying that a significant fraction of Pdx-1 and NeuroD1 binding sites exist within regulatory modules, where they may control gene expression in a synergistic manner. However, it was unknown whether MafA interacts at similar loci to co-regulate genes important for β-cell function, as no genome-wide ChIP binding studies have been performed for MafA. I have shown in Chapter 3 that Pdx1, NeuroD1 and MafA bind to one another's promoters, along with the *Insulin* promoter. The Pdx1, NeuroD1 and MafA peaks are also enriched around all three peak centers, suggesting they associated strongly with one another. There was a large number of Pdx1-only peaks and a big overlap of NeuroD1 and MafA binding sites, and this similarity was also reflected in the motifs identified for MafA and NeuroD1. This suggests that Pdx1 is acting more independently, with NeuroD1 and MafA having very similar and possibly redundant roles. A more independent role of Pdx1 would also concur with the fact that 17% of Pdx1 peaks are found in promoter regions, compared to NeuroD1 and MafA. Depletion of NeuroD1 has more

effect on the target genes it binds to, 834 genes are regulated by NeuroD1 only, in comparison to the 175 and 151 regulated by Pdx1 and MafA respectively. This suggests that NeuroD1 is acting in an indirect manner. In all three cases, there are a greater number of genes being up-regulated upon knock-down, indicating that these transcription factors have a generally more repressive role on their target genes, rather than activating. Gene ontology analysis on the target genes revealed over-representation in categories such as metabolic processes, developmental processes and cellular processes.

As the binding sites of Pdx1, NeuroD1 and MafA and the genes they regulated categorized into similar regulatory networks, and these three transcription factors synergistically bind to and regulate the *Insulin* promoter[124], I hypothesized that these three transcriptional factors would combinatorially regulate other genes in the genome. To explore how combinatorial transcription factor binding correlated with changes in gene expression upon combinatorial transcription factor knockdown, the number of regulated genes that had any combination of a transcription factor was overlapped with the genes identified to be changing upon each combination of transcription factor knockdown. Genes containing Pdx1 binding sites in their promoters display maximal differential expression in the single, double or triple knockdown experiments (Figure 28). The highest percentage of genes regulated are those in the double knockdown, irrespective of whether the gene is bound by single, double or all three transcription factors (Figure 29). Of the single transcription factor regulated genes, NeuroD1-regulated genes have a higher percentage of bound genes (6-8%) compared to Pdx1 and MafA (<4%). This could be due to the fact that 12.5% of NeuroD1-regulated genes are transcription factors (Table 6), as compared to 8.5% of Pdx1- and MafA- regulated genes. This higher proportion of transcription factors in NeuroD1 regulated genes again suggests that NeuroD1 mediates downstream gene expression changes via other transcription factors. Taken together, I have shown that while Pdx1 on its own had the most effect on gene regulation, triply and doubly bound genes exert more effect on gene regulation.

7.2. PDX1, NEUROD1 AND MAFA TARGET GENES HAVE FUNCTIONAL EFFECTS ON THE BETA-CELL PHENOTYPE

Transcription factors play an important role in pancreatic β-cell development and function. This is manifested clinically by MODY syndromes, in which heterozygous mutations in HNF4α, HNF1α, PDX1, HNF1β and NEUROD1 cause progressive impairments in insulin secretion and the eventual development of diabetes mellitus[90]. The main function of the mature β-cell is to respond appropriately to the fluctuating levels of blood glucose with the secretion of insulin. This is mediated and controlled by many transcription factors, of which Pdx1, NeuroD1 and MafA are a few. As Pdx1, NeuroD1 and MafA have been shown to co-regulate many genes of similar biological functions (Chapter 3), I hypothesized that these Pdx1 NeuroD1 and MafA target genes were important in maintaining the insulin secretory response, and that depletion of their expression in β-cells would result in impaired ability to respond to glucose, as determined by a GSIS assay. In Chapter 4, I have shown that Pdx1, NeuroD1 and MafA repress *Itgb1bp2* and *Cplx2* in order to achieve an enhanced GSIS response. Pdx1, NeuroD1 and MafA activate *Tspyl1*, *F13a1* and *Septin7* to reduce GSIS (Figure 39). This additional knowledge emphasizes the exquisite level of regulation required for the appropriate response of β-cells to a glucose stimulus.

Future work involves extending the study to more genes regulated by these three transcription factors, in order to identify new regulators important in GSIS. The β-cell is also responsive to other nutrients and numerous neural and hormonal factors, and specific amino acids may acutely and/or chronically regulate insulin secretion[347]. Therefore it is also of biological interest to study the role of these commonly regulated target genes in response to different metabolic stimulus, as they might be also involved in different aspects of β-cell function.

7.3. CHROMATIN REMODELLING IN RESPONSE TO BETA-CELL DYSFUNCTION

The discovery of defects in epigenetic modification in T1DM in susceptibility loci makes it one of the strongest candidates for that elusive "missing link" of auto-immune disease research, providing a mechanism common to and potentially linking both environment stimulation and genetic susceptibility in immunopathogenesis. The activation of β-cell genes in the pancreas requires an appropriate euchromatin DNA template characterized by hyper-methylation of H3K4 that is maintained by the HMT Set7/9[162].Therefore, I hypothesized that there would be chromatin remodelling in β-cells in response to a glucolipotoxic stimulus that impairs its ability to secrete insulin in response to glucose. In Chapter 5, I profiled marks of active and repressive chromatin in normal and dysfunctional β-cells, and found that there was a global increase of chromatin opening in dysfunctional β-cells, as indicated by the increase in numbers of FAIRE and H3K4me3 peaks at the promoters of genes, and a decrease in H3K27me3 at the promoters of genes in dysfunctional β-cells. Motif analysis of FAIRE peaks in dysfunctional β-cells showed that there was a bias for binding sites for general transcription factors e.g. SP-1, CTCF, NF-Y, while motifs for FAIRE peaks in normal β-cells showed enrichments for tissue-specific transcription factors like FoxA1 and FoxA2, Oct2 and Oct4, and Pdx1. Comparison of the genes associated with the respective epigenetic marks in normal and dysfunctional β-cells demonstrated that there is a significant increase in genes involved in cellular regulatory activities and the specification of other developmental lineages in dysfunctional β-cells.

Genes important for β-cell function are marked by H3K4me3 in their promoters in normal β-cells, e.g. *Pdx1, NeuroD1, MafA, Ins1, Glut2, Snap25, Vamp2, Cacna1a, Chgb, Kirrel6.2, Sytl4, Stxbp1,* and *Syt7.* On a global scale, genes bound by Pdx1, NeuroD1 and MafA in normal β-cells are associated with H3K4me3, not H3K27me3 (Figure 50). There is an increase in the association of these transcription factors in dysfunctional β-cells with the repressive H3K27me3 mark, and a corresponding decrease in gene expression (Figure 52).

Taking into account there was reduced expression of *Pdx1*, *NeuroD1* and *MafA* in dysfunctional β-cells, the H3K27me3-marked genes (that have a TFBS) in dysfunctional β-cells are enriched for genes involved in metabolic processes, developmental processes and general cellular regulatory processes (e.g. cell cycle, cell signalling). I therefore propose a model in which glucolipotoxic conditions lead to increase in ROS and other metabolic changes that result in a reduced H3K4me3 mark of *Pdx1*, *NeuroD1* and *MafA* and correspondingly lower mRNA levels in dysfunctional β-cells. This in turn leads to chromatin remodelling at the promoters of their target genes, resulting in changes in gene expression that ultimately affect β-cell function. More work is needed (highlighted in Chapter 5.3.2.) to truly understand the molecular mechanisms behind chromatin remodelling in β-cell dysfunction.

Bivalent genes marked by both active and repressive chromatin marks, though first identified in ESCs, are not ESC-specific[179,326]. In this thesis, I show that these apparently conflicting domains exist in normal β-cells and majority of them lose their H3K27me3 marks in dysfunctional β-cells, consistent with the global increase in the numbers of H3K4me3 and FAIRE peaks and decrease in the global numbers of H3K27me3 peaks in dysfunctional β-cells. Genes that are bivalently marked in normal β-cells are more enriched for in developmental processes and cell regulatory processes, compared to genes that are bivalently marked in normal β-cells and have lost the repressive H3K27me3 mark in dysfunctional β-cells. Some examples of bivalent genes in normal β-cells that have reduced levels of H3K4me3 and/or increased levels of H3K27me3 in dysfunctional β-cells include genes involved in pancreatic development and/or function, e.g. *MafB, Oxr1, Camk2b, miR-9, Sox17;* genes involved in neuronal development and function, e.g. *Sema3d* and *Celsr1*; members of the *Fox* family transcription factors e.g. *Foxf2* and *Foxq1*; and the potassium channels *Kcnc1* and *Kcnq1* (Figure 56). Therefore, I have shown that bivalent chromatin states are modified to different extents in dysfunctional β-cells and result in gene expression changes that could ultimately contribute to its inability to properly respond to glucose.

The interplay between epigenetic and transcriptional regulation is a complex process that depends on numerous factors. Transitions between different chromatin states are dynamic and depend on a balance between factors that promote a repressive state and those than promote a transcriptionally active one. Therefore, drugs that target different components of the epigenetic machinery can co-operate in restoring normal gene activity. For example, 5-aza-2'-deoxycytidine interferes with the activity of Dnmt1, leading to genome hypo-methylation, reactivating tumor suppressor genes that were silenced[348]. However, it is toxic, and cannot be used for therapeutic purposes on its own[349]. When combined with the HDAC inhibitor TSA, it reactivates silenced tumor suppressor genes in cancer cells[350]. This suggests the combination of chromatin modifiers can be used to treat diseases.

Clinical trials with HDAC inhibitors have repeatedly demonstrated low toxicity despite the expectation that they may cause global gene dysregulation, and preliminary studies indicate that they may prove useful in treating diabetes[185]. The HDAC Sirt1 stimulates several important signaling pathways involved in regulating energy utilization, including the synthesis of new mitochondria[351]. Mitochondrial activity in metabolically active tissues (such as muscle) will increase metabolic rate, drive glucose metabolism and fatty acid oxidation and thereby improve insulin sensitivity and enhance energy expenditure in multiple tissues[352,353]. Sirt1 activation also plays a significant role on other key cellular regulators involved in metabolic and oxidative stress. In a DIO mouse model, new chemical entities (NCEs) at Sirtris Pharmaceuticals[354] significantly improved insulin sensitivity and glucose levels in a similar manner observed in DIO mice genetically altered to have increased *Sirt1* levels[355]. DIO mice treated with these same NCE SIRT1 activators also had reduced weight gain, increased energy expenditure and had global profiles of metabolites in tissues and plasma that were consistent with a reversal of the metabolic dysfunction induced by the high fat diet[356].

It is desirable that the balance of genes whose regulation would be altered with by epigenetic modifiers would tip the balance from unfavorable diabetogenes to those with enhanced benefits. Results I have presented in my thesis have identified areas of chromatin remodeling in β-cell dysfunction. With the insights gained from that and future experiments that look at epigenomic changes in animal models of diabetes or diabetic patient islets, coupled with screening for drugs that can modulate the chromatin environment to reverse the dysfunctional β-cell phenotype, it is hoped that eventually, genome-wide transcriptional and epigenetic characterization will lead to patient-specific therapies that also allow monitoring of genome-wide epigenetic consequences of these therapies.

7.4. CONCLUDING REMARKS

This thesis emphasizes a role of transcriptional regulation in the maintenance of pancreatic β-cell function, and presents unique evidence for chromatin remodelling in β-cell dysfunction. The genetic information of an organism is differentially regulated in a spatial and temporal manner through complex mechanisms that we are only just beginning to comprehend. Epigenetic modifications control gene expression by changing specific regions of the genome to maintain silencing or activating activity, often in response to an environmental stimulus. The trigger for the differential marking of the genome remains largely uncharacterized, but remarkable progress is being made on this front. In order to fully harness the potential of chromatin-modifying drugs in complex diseases like Type 2 Diabetes, it is imperative to better understand the molecular mechanisms by which transcription factors interact with the chromatin, and the impact of the environment on these interactions. To this end, my work presents novel insights into Pdx1-, NeuroD1- and MafA-regulated pathways that influence β-cell function, and the effect of glucolipotoxicity on chromatin remodeling in β-cell dysfunction.

Symbol	Definition
ATP	Adenosine triphosphate
bHLH	basic helix-loop-helix
bLZ	Basic leucine-leucine zipper
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chomatin immunoprecipitation-chip
ChIP-Seq	ChIP coupled with high-throughput sequencing
CHO	Chinese hamster ovary
COREs	Clusters of open regulatory elements
DMEM	Dulbecco modified Eagle's medium
DIO	Diet-induced obese
ECM	Extra-cellular matrix
ER	Endoplasmic reticulum
ESC	Embryonic stem cells
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FFA	Free fatty acids
GSIS	Glucose stimulated insulin secretion
GWAS	Genome-wide association studies
HD	Homeodomain
HDL-C	High-density lipoprotein cholesterol
IDDM	Insulin dependent diabetes mellitus
IEGs	Immediate-early genes
IPA	Ingenuity Pathway Analysis
IUGR	Intrauterine growth retardation
IVG	Integrated Genomics Viewer
KRH	Krebs-Ringers-Hepes
LDL-C	Low-density lipoprotein cholesterol
LIF	Leukemia inhibitory factor
MEFs	Mouse embryonic fibroblasts
miRNA	MicroRNA
MODY	Maturity Onset Diabetes of the Young
NADH	Nicotinamide dinucleotide
nc	Non-coding
NTC	Non-targeting control
PBS	Phosphate buffered saline
RA	Retinoic acid
RGS	Regulators of G protein signaling
RNAi	RNA interference
ROS	Reactive oxygen species
SACO	Serial analysis of chromatin occupancy
SNPs	Single nucleotide polymorphisms
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
TFBS	Transcription factor binding site
TSA	Trichostatin A
TSS	Transcription start site
UCSC	University of California, Santa Cruz
ZDF	Zucker Diabetic Fatty

GENE ABBREVIATIONS

Gene symbol	Gene name
Acac	Acetyl-CoA carboxylase
Aldob	Aldolase B fructose-bisphosphate
Ampk	Protein kinase AMP-activated β1 non-catalytic subunit
Arnt2	Aryl-hydrocarbon receptor nuclear translocator 2
Arx	Aristaless related homeobox
Atp1b1	ATPase Na+/K+ transporting β1 polypeptide
Bend	BEN domain containing
BMP	Bone morphogenic protein
Cacna	Calcium channel voltage-dependent P/Q type
Camk	Calcium/calmodulin-dependent protein kinase
cAMP	Cyclic adenosine monophosphate
Celsr	Cadherin EGF LAG seven-pass G-type receptor
Chgb	Chromogranin B
Clic	Chloride intracellular channel
Cplx	Complexin
Cre	cAMP-response element
CTCF	CCCTC-binding factor
DNMT	DNA methyltransferase
Elfn	Extracellular leucine-rich repeat and fibronectin type III domain containing
Exoc	Exocyst complex component
F13a1	Coagulation factor XIII A1 polypeptide
FGF	Fibroblast growth factor
Fox	Forkhead box
Gad	Glutamate decarboxylase
Gcg	Glucagon
Gcnt	Glucosaminyl (N-acetyl) transferase 2, I-branching enzyme (I blood group)
Gk	Glucokinase
Glp	Glucagon-like peptide
Glut	Glucose transporter
GPDH	Glycerol phosphate dehydrogenase
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
Hist1	Histone cluster 1
HMT	Histone methyltransferase
Hnf	Hepatocyte nuclear factor
Hox	Homeobox
IA1	Insulinoma-associated 1
Iapp	Islet amyloid polypeptide
IGF	Insulin-like growth factor
IL	Interleukin
IRS	Insulin receptor substrate
Itg1bp	Integrin β 1 binding protein (melusin)
JNK	Mitogen-activated protein kinase 8
Kcnc1	Potassium voltage-gated channel Shaw-related subfamily member 1
Kcnq1	Potassium voltage-gated channel, KQT-like subfamily, member 1
Kirrel 2	Kin of IRRE like 2
Kirrel6.2	Potassium inwardly-rectifying channel subfamily J member 11
M3R	M3 muscarinic receptors
MafA	v-maf musculoaponeurotic fibrosarcoma oncogene homolog
MeCP	Methyl-CpG-binding protein
Nanog	Nanog homeobox

Gene symbol	Gene name
Ndrg	NDRG family member
NeuroD1	Neurogenic differentiation 1
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
Ngn3	Neurogenin 3
Nkx	NK homeobox
Nna	Neuronatin
Oct2	POU class 3 homeobox 2
Oct4	POU class 5 homeobox 1
Oxa	Orexin A
Oxr	Oxidation resistance
Pax	Paired-box
Pcsk1	Proprotein convertase subtilisin/kexin type 1
Pdx1	Pancreatic duodenal homeobox 1
PEPCK	Phosphoenolpyruvate carboxykinase
Ppp1r	Protein phosphatase 1, regulatory (inhibitor) subunit
PPY	Pancreatic polypeptide
Ptprn	Protein tyrosine phosphatase receptor type N
Sema	Semaphorin
Serca	ATPase, Ca^{2+} transporting cardiac muscle fast twitch 1
Shc	Src homology 2 domain containing transforming protein
Shh	Sonic hedgehog
SHP	Small heterodimer partner
Sik	Salt-inducible kinase
Sirt	Sirtuin
Slc1a2	Solute carrier family 1 (glial high affinity glutamate transporter) member 2
Snap	Synaptosomal-associated protein
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
Sox	SRY (sex determining region Y)-box
Sp1	Sp transcription factor 1
Srebp	Sterol regulatory element binding protein
Sst	Somatostatin
Stx	Syntaxin
Sur	Sulfonylurea receptor
Syt	Synaptogamin
Tcf7l2	Transcription factor 7–like 2
TGF-β	Transforming growth factor β
Tmem27	Transmembrane 27
TNF-α	Tumor necrosis factor α
Tspyl	TSPY-like
TTR	Transthyretin
Ucp2	Uncoupling protein 2
Vamp	Vesicle-associated membrane protein
WFS1	Wolfram syndrome 1

BIBLIOGRAPHY

1. Malecki M.T., et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nature Genetics* **23**, 323-328 (1999).

2. Lawless, M. W., et al. Targeting histone deacetylases for the treatment of disease: Epigenetics Review Series. *Journal of Cellular and Molecular Medicine* **13**, 826-852 (2009).

3. World Health Organization & International Diabetes Federation *Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia*. 1-41 (2006).

4. Ren, J. et al. Pancreatic islet cell therapy for type I diabetes: understanding the effects of glucose stimulation on islets in order to produce better islets for transplantation. *Journal of translational medicine* **5**, 1 (2007).

5. Prasad, M. et al. Association between type 1 diabetes and GWAS SNPs in the southeast US Caucasian population. *Genes and Immunity* **12**, 208-212 (2011).

6. Martin B.C., et al. Role of glucose and insulin resistance in development of type 2 diabetes mellitus: Results of a 25-year follow-up study. *Lancet* **340**, 925-929 (1992).

7. Porte, D. Clinical importance of insulin secretion and its interaction with insulin resistance in the treatment of type 2 diabetes mellitus and its complications. *Diabetes/Metabolism Research and Reviews* **17**, 181-188 (2001).

8. Doria, A., Patti, M.-E. & Kahn, C.R. The emerging genetic architecture of type 2 diabetes. *Cell metabolism* **8**, 186-200 (2008).

9. Weyer C., et al. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *Journal of Clinical Investigation* **104**, 787-794 (1999).

10. Mozaffarian, D., et al. "Lifestyle risk factors and new-onset diabetes mellitus in older adults: the cardiovascular health study". *Archives of Internal Medicine* **169**, 798-807 (2009).

11. Centers for Disease Control and Prevention Prevalence of overweight and obesity among adults with diagnosed diabetes—United States, 1988–1994 and 1999–2002. *Morbidity and Mortality Weekly Report* **53**, 1066-8 (2004).

12. Risérus, U., et al. Dietary fats and prevention of type 2 diabetes. *Progress in Lipid Research* **48**, 48-51 (2009).

13. Zhang, Y., et al. Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425-32 (1994).

14. Das, A. & Mukhopadhyay, S. The evil axis of obesity, inflammation and type-2 diabetes. *Endocrine, metabolic & immune disorders drug targets* **11**, 23-31 (2011).

15. Lang, I.A., et al. Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults. *JAMA* **300**, 1303-10 (2008).

16. Steffes, M.W., et al. β-cell function and the development of diabetes-related complications in the diabetes control and complications trial. *Diabetes Care* **26**, 832-836 (2003).

17. Ripsin, C., et al. Management of blood glucose in type 2 diabetes mellitus. *Am Fam Physician* **79**, 29-36 (2009).

18. Kirpichnikov, D., et al. Metformin: an update. *Ann Intern Med.* **137**, 25-33 (2002).

19. Chang, A.M., Halter, J. Aging and insulin secretion. *American Journal of Physiology. Endocrinology and Metabolism* **284**, E7-12 (2003).

20. Shapiro A.M., et al. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *New England Journal of Medicine* **343**, 230-238 (2000).

21. Shapiro A.M., et al. International trial of the Edmonton protocol for islet transplantation. *New England Journal of Medicine* **355**, 1318-1330 (2006).

22. Hirshberg B., et al. Benefits and Risks of Solitary Islet Transplantation for Type 1 Diabetes Using Steroid-Sparing Immunosuppression: The National Institutes of Health experience. *Diabetes Care* **26**, 3288-3295 (2003).

23. Dor, Y., et al. a Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41-6 (2004).

24. Xu, X. et al. Beta cells can be generated from endogenous progenitors in injured adult mouse pancreas. *Cell* **132**, 197-207 (2008).

25. Smukler, S. et al. The Adult Mouse and Human Pancreas Contain Rare Multipotent Stem Cells that Express Insulin. *Cell stem cell* **8**, 281-293 (2011).

26. Stanley, E.G. & Elefanty, A.G. Building better beta cells. *Cell stem cell* **2**, 300-1 (2008).

27. Kroon, E. et al. Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nature biotechnology* **26**, 443-52 (2008).

28. Zhou, Q., et al. a In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627-32 (2008).

29. Hering B.J., et al. Prolonged diabetes reversal after intraportal xenotransplantation of wild-type porcine islets in immunosuppressed nonhuman primates. *Nature Medicine* **12**, 301-303 (2006).

30. Murtaugh, L.C. & Melton, D. a Genes, signals, and lineages in pancreas development. *Annual review of cell and developmental biology* **19**, 71-89 (2003).

31. Deltour L., et al. Polyclonal origin of pancreatic islets in aggregation mouse chimaeras. *Development* **112**, 1115-1121 (1991).

32. Steiner, D.J., et al. Interspecies comparison of islet architecture and composition. *Endocrine* 135-145 (2010).

33. Cabrera O., et al. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 2334-2339 (2006).

34. Kim A., et al. Islet architecture: A comparative study. *Islets* **1**, 129-136 (2009).

35. Weir, G.C. & Bonner-weir, S. Five stages of evolving beta cell dysfunction during progression to diabetes. *Diabetes* **53**, S16-21 (2004).

36. Thorens B., et al. Molecular physiology of glucose transporters. *Diabetes Care* **13**, 209-218 (1990).

37. MacDonald M.J. et al. Lack of glyconeogenesis in pancreatic islets: Expression of gluconeogenic enzyme genes in islets. *Hormone and Metabolic Research* **24**, 158-160 (1992).

38. Sekine N., et al. Low lactate dehydrogenase and high mitochondrial glycerol phosphate dehydrogenase in pancreatic β-cells. Potential role in nutrient sensing. *Journal of Biological Chemistry* **269**, 4895-4902 (1994).

39. MacDonald, M.J. Elusive proximal signals of β-cells for insulin secretion. *Diabetes* **39**, 1461-1466 (1990).

40. MacDonald, M.J. Feasibility of a mitochondrial pyruvate malate shuttle in pancreatic islets. Further implication of cytosolic NADPH in insulin secretion. *Journal of Biological Chemistry* **270**, 20051-20058 (1995).

41. Hoogwerf, B.J., Goetz, C. Urinary C-peptide: a simple measure of integrated insulin production with emphasis on the effects of body size, diet, and corticosteroids. *Jour Clin Endocrinol Metab* **56**, 60-7 (1983).

42. Ido, Y., et al. Prevention of vascular and neural dysfunction in diabetic rats by C-peptide. *Science* **277**, 563-6 (1997).

43. Pittner, R.A., et al. Molecular physiology of amylin. *J. Cell. Biochem.* **55**, 19-28 (1994).

44. AS, F. et al. *Harrison's Principle of Internal Medicine*. Insulin Biosynthesis (2008).

45. Paolisso G.,et al. A high concentration of fasting plasma non-esterified fatty acids is a risk factor for the development of NIDDM. *Diabetologia* **38**, 1213-1217 (1995).

46. Chen C., et al. Mechanism of compensatory hyperinsulinemia in normoglycemic insulin- resistant spontaneously hypertensive rats. Augmented enzymatic activity of glucokinase in β-cells. *Journal of Clinical Investigation* **94**, 399-404 (1994).

47. Ahrén, B. Autonomic regulation of islet hormone secretion - Implications for health and disease. *Diabetologia* **43**, 393-410 (2000).

48. Prentki M., et al. Islet β cell failure in type 2 diabetes. *Journal of Clinical Investigation* **116**, 1802-1812 (2006).

49. Tokuyama Y., et al. Evolution of β-cell dysfunction in the male Zucker diabetic fatty rat. *Diabetes* **44**, 1447-1457 (1995).

50. Laybutt, D.R. et al. Genetic regulation of metabolic pathways in beta-cells disrupted by hyperglycemia. *The Journal of biological chemistry* **277**, 10912-21 (2002).

51. Quintens, R., et al. Why expression of some genes is disallowed in beta-cells. *Biochem Soc Trans* **3**, 300-5 (2008).

52. Laybutt, D.R., et al. Increased expression of antioxidant and antiapoptotic genes in islets that may contribute to beta-cell survival during chronic hyperglycemia. *Diabetes* **51**, 413-23 (2002).

53. Maedler K., et al. Glucose-induced β cell production of IL-1β contributes to glucotoxicity in human pancreatic islets. *Journal of Clinical Investigation* **110**, 851-860 (2002).

54. Kilpatrick, et al. Differentiation between glucose-induced desensitization of insulin secretion and β-cell exhaustion in the HIT-T15 cell line. *Diabetes* **47**, 606-611 (1998).

55. Vague P., et al. The defective glucose sensitivity of the B cell in non insulin dependent diabetes. Improvement after twenty hours of normoglycaemia. *Metabolism: Clinical and Experimental* **31**, 139-142 (1982).

56. Poitout V., et al. Glucolipotoxicity: Fuel excess and β-cell dysfunction. *Endocrine Reviews* **29**, 351-366 (2008).

57. Jacqueminet S., et al. Inhibition of insulin gene expression by long-term exposure of pancreatic β cells to palmitate is dependent on the presence of a stimulatory glucose concentration. *Metabolism: Clinical and Experimental* **49**, 532-536 (2000).

58. Olson L.K., et al. Reduction of insulin gene transcription in HIT-T15 β cells chronically exposed to a supraphysiologic glucose concentration is associated with loss of STF-1 transcription factor expression. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9127-9131 (1995).

59. Harmon J.S., et al. Oxidative stress-mediated, post-translational loss of MafA protein as a contributing mechanism to loss of insulin gene expression in glucotoxic beta cells. *Journal of Biological Chemistry* **280**, 11107-11113 (2005).

60. Robertson, R.P. Beta-cell deterioration during diabetes: what's in the gun? *Trends in endocrinology and metabolism: TEM* **20**, 388-93 (2009).

61. Lee C.S., et al. Foxa2 controls Pdx1 gene expression in pancreatic β-cells in vivo. *Diabetes* **51**, 2546-2551 (2002).

62. Lantz , K.A., et al. Foxa2 regulates multiple pathways of insulin secretion. *Journal of Clinical Investigation* **114**, 512-520 (2004).

63. Jensen J., et al. Control of endodermal endocrine development by Hes-1. *Nature Genetics* **24**, 36-44 (2000).

64. Shih D.Q., et al. Loss of HNF-1α Function in Mice Leads to Abnormal Expression of Genes Involved in Pancreatic Islet Development and Metabolism. *Diabetes* **50**, 2472-2480 (2001).

65. Hansen S.K., et al. Genetic evidence that HNF-1α-dependent transcriptional control of HNF-4α is essential for human pancreatic β cell function. *Journal of Clinical Investigation* **110**, 827-833 (2002).

66. Roth U., et al. The Transcription Factors HIF-1 and HNF-4 and the Coactivator p300 Are Involved in Insulin-regulated Glucokinase Gene Expression via the Phosphatidylinositol 3-Kinase/Protein Kinase B Pathway. *Journal of Biological Chemistry* **279**, 2623-2631 (2004).

67. Thomas H., et al. A distant upstream promoter of the HNF-4α gene connects the transcription factors involved in maturity-onset diabetes of the young. *Human Molecular Genetics* **10**, 2089-2097 (2001).

68. Ahlgren U., et al. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature* **385**, 257-260 (1997).

69. Zhang C., et al. MafA is a key regulator of glucose-stimulated insulin secretion. *Molecular and Cellular Biology* **25**, 4969-4976 (2005).

70. Nishimura, W., et al. A switch from MafB to MafA expression accompanies differentiation to pancreatic β-cells. *Developmental Biology* **293**, 526-539 (2006).

71. Naya, F.J. et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice. *Genes and Development* **11**, 2323-2334 (1997).

72. Watada, H. Neurogenin 3 is a key transcription factor for differentiation of the endocrine pancreas. *Endocrine Journal* **51**, 255-264 (2004).

73. Sander N., et al. Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of β-cell formation in the pancreas. *Development* **127**, 5533-5540 (2000).

74. Sussel, L., et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic β cells. *Development* **125**, 2213-2221 (1998).

75. Schisler, J.C. et al. The Nkx6.1 homeodomain transcription factor suppresses glucagon expression and regulates glucose-stimulated insulin secretion in islet beta cells. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7297-302 (2005).

76. Ritz-Laser B., et al.The pancreatic beta-cell-specific transcription factor Pax-4 inhibits glucagon gene expression through Pax-6. *Diabetologia* **45**, 97-107 (2002).

77. Ritz-Laser B., et al. Pax-6 and Cdx-2/3 interact to activate glucagon gene expression on the G1 control element. *Journal of Biological Chemistry* **274**, 4124-4132 (1999).

78. Oliver-krasinski, J.M. et al. The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice. doi:10.1172/JCI37028.The

79. Offield M.F., et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983-995 (1996).

80. Mosley A.L., et al. Glucose regulation of insulin gene expression requires the recruitment of p300 by the β-cell-specific transcription factor Pdx-1. *Molecular Endocrinology* **18**, 2279-2290 (2004).

81. Ahlgren, U. et al. β-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the β-cell phenotype and maturity onset diabetes. *Genes and Development* **12**, 1763-1768 (1998).

82. Wells J.M., et al. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* **127**, 1563-1572 (2000).

83. St-Onge L., et al. Pancreas development and diabetes. *Current Opinion in Genetics and Development* **9**, 295-300 (1999).

84. Jørgensen, M.C. et al. An illustrated review of early pancreas development in the mouse. *Endocrine reviews* **28**, 685-705 (2007).

85. Chakrabarti S.K., et al. Transcription factors direct the development and function of pancreatic beta cells. *Trends in endocrinology and metabolism: TEM* **14**, 78-84 (2003).

86. Keller, D.M. et al. Characterization of pancreatic transcription factor Pdx-1 binding sites using promoter microarray and serial analysis of chromatin occupancy. *The Journal of biological chemistry* **282**, 32084-92 (2007).

87. Collombat P., et al. The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the α- and β-cell lineages in the mouse endocrine pancreas. *Development* **132**, 2969-2980 (2005).

88. Ashery-Padan R., et al. Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes. *Developmental Biology* **269**, 479-488 (2004).

89. Murtaugh, L.C. Pancreas and beta-cell development : from the actual to the possible. *Development* **438**, 427-438 (2007).

90. Fajans S.S., et al. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *New England Journal of Medicine* **345**, 971-980 (2001).

91. Waeber G., et al.Transcriptional activation of the GLUT2 gene by the IPF-1/STF-1/IDX-1 homeobox factor. *Molecular Endocrinology* **10**, 1327-1334 (1996).

92. Watada H., et al.The human glucokinase gene β-cell-type promoter: An essential role of insulin promoter factor 1/PDX-1 in its activation in HIT-T15 cells. *Diabetes* **45**, 1478-1488 (1996).

93. Habener J.F., et al. Minireview: Transcriptional regulation in pancreatic development. *Endocrinology* **146**, 1025-1034 (2005).

94. Pedersen J.K., et al. Endodermal expression of Nkx6 genes depends differentially on Pdx1. *Developmental Biology* **288**, 487-501 (2005).

95. Jonsson J., et al. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606-609 (1994).

96. Iype T., et al. Mechanism of insulin gene regulation by the pancreatic transcription factor Pdx-1: Application of pre-mRNA analysis and chromatin immunoprecipitation to assess formation of functional transcriptional complexes. *Journal of Biological Chemistry* **280**, 16798-16807 (2005).

97. Qiu Y., et al. Insulin gene transcription is mediated by interactions between the p300 coactivator and PDX-1, BETA2, and E47. *Molecular and Cellular Biology* **22**, 412-420 (2002).

98. Chakrabarti, S.K. et al. Covalent histone modifications underlie the developmental regulation of insulin gene transcription in pancreatic beta cells. *The Journal of biological chemistry* **278**, 23617-23 (2003).

99. Samaras S.E., et al. Conserved sequences in a tissue-specific regulatory region of the pdx-1 gene mediate transcription in pancreatic β cells: Role for hepatocyte nuclear factor 3β and Pax6. *Molecular and Cellular Biology* **22**, 4702-4713 (2002).

100. Marshak S., et al. Functional conservation of regulatory elements in the pdx-1 gene: PDX-1 and hepatocyte nuclear factor 3β transcription factors mediate β-cell-specific expression. *Molecular and Cellular Biology* **20**, 7583-7590 (2000).

101. Sharma S., et al. Hormonal regulation of an islet-specific enhancer in the pancreatic homeobox gene STF-1. *Molecular and Cellular Biology* **17**, 2598-2604 (1997).

102. Ben-Shushan E., et al. A Pancreatic β-Cell-specific Enhancer in the Human PDX-1 Gene Is Regulated by Hepatocyte Nuclear Factor 3β (HNF-3β), HNF-1α, and SPs Transcription Factors. *Journal of Biological Chemistry* **276**, 17533-17540 (2001).

103. Kitamura T., et al. The forkhead transcription factor Foxo1 links insulin signaling to Pdx1 regulation of pancreatic β cell growth. *Journal of Clinical Investigation* **110**, 1839-1847 (2002).

104. Gray, S.G. & De Meyts, P. Role of histone and transcription factor acetylation in diabetes pathogenesis. *Diabetes/metabolism research and reviews* **21**, 416-33 (2005).

105. Lee, J.E. NeuroD and neurogenesis. *Developmental Neuroscience* **19**, 27-32 (1997).

106. Mutoh H., et al. The basic helix-loop-helix transcription factor BETA2/neuroD is expressed in mammalian enteroendocrine cells and activates secretin gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3560-3564 (1997).

107. Glick E., et al. Transcription factor BETA2 acts cooperatively with E2A and PDX1 to activate the insulin gene promoter. *Journal of Biological Chemistry* **275**, 2199-2204 (2000).

108. Dumonteil, E., et al. Differential regulation of the glucagon and insulin I gene promoters by the basic helix-loop-helix transcription factors E47 and BETA2. *Journal of Biological Chemistry* **273**, 19945-19954 (1998).

109. Moates J.M., et al. BETA2 activates transcription from the upstream glucokinase gene promoter in islet β-cells and gut endocrine cells. *Diabetes* **52**, 403-408 (2003).

110. Martin, C.C. et al. Upstream stimulatory factor (USF) and neurogenic differentiation/β-cell E box transactivator 2 (NeuroD/BETA2) contribute to islet-specific glucose-6-phosphatase catalytic-subunit-related protein (IGRP) gene expression. *Journal of Biochemistry* **371**, 675–686 (2003).

111. Kim J.-W., et al. Transactivation of the mouse sulfonylurea receptor I gene by BETA2/NeuroD. *Molecular Endocrinology* **16**, 1097-1107 (2002).

112. Itkin-Ansari P., et al. NeuroD1 in the endocrine pancreas: Localization and dual function as an activator and repressor. *Developmental Dynamics* **233**, 946-953 (2005).

113. Chu, K. & Tsai, M.J. Neuronatin, a downstream target of BETA2/NeuroD1 in the pancreas, is involved in glucose-mediated insulin secretion. *Diabetes* **54**, 1064–1073 (2005).

114. Ishizuka, N., et al. Induction by NeuroD of the components required for regulated exocytosis. *Biochemical and Biophysical Research Communications* **354**, 271-77 (2007).

115. Shieh S.-Y., et al. Cell-specific and ubiquitous factors are responsible for the enhancer activity of the rat insulin II gene. *Journal of Biological Chemistry* **266**, 16708-16714 (1991).

116. Olbrot M., et al. Identification of β-cell-specific insulin gene transcription factor RIPE3b1 as mammalian MafA. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6737-6742 (2002).

117. Samaras S.E., et al. The islet β cell-enriched RIPE3b1/Maf transcription factor regulates pdx-1 expression. *Journal of Biological Chemistry* **278**, 12263-12270 (2003).

118. Nishizawa M., et al. v-maf, a viral oncogene that encodes a "leucine zipper" motif. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 7711-7715 (1989).

119. Ogino H., et al. Induction of lens differentiation by activation of a bZIP transcription factor, L-Maf. *Science* **280**, 115-118 (1998).

120. Kitamura Y.I., et al. FoxO1 protects against pancreatic β cell failure through NeuroD and MafA induction. *Cell Metabolism* **2**, 153-163 (2005).

121. Raum, J.C. et al. FoxA2, Nkx2.2, and PDX-1 regulate islet β-cell-specific mafA expression through conserved sequences located between base pairs -8118 and -7750 upstream from the transcription start site. *Molecular and Cellular Biology* **26**, 5735-5743 (2006).

122. Vanhoose, A. et al. MafA and MafB Regulate Pdx1 Transcription through the Area II Control Region in Pancreatic β Cells. *Journal of Biological Chemistry* **283**, 22612–22619 (2008).

123. Kaneto H., et al. PDX-1/VP16 fusion protein, together with neurod or Ngn3, markedly induces insulin gene transcription and ameliorates glucose tolerance. *Diabetes* **54**, 1009-1022 (2005).

124. Aramata, S., et al. Synergistic activation of the insulin gene promoter by the beta-cell enriched transcription factors MafA, Beta2, and Pdx1. *Biochimica et biophysica acta* **1730**, 41-6 (2005).

125. Stoffers D.A., et al. Insulinotropic glucagon-like peptide 1 agonists stimulate expression of homeodomain protein IDX-1 and increase islet size in mouse pancreas. *Diabetes* **49**, 741-748 (2000).

126. Cissell M.A., et al. Transcription factor occupancy of the insulin gene in vivo: Evidence for direct regulation by Nkx2.2. *Journal of Biological Chemistry* **278**, 751-756 (2003).

127. Zhao L., et al. The islet β cell-enriched MafA activator is a key regulator of insulin gene transcription. *Journal of Biological Chemistry* **280**, 11887-11894 (2005).

128. Barrow, J., et al. Transcription factor cycling on the insulin promoter. *FEBS letters* **580**, 711-5 (2006).

129. McCarthy, M.I. Genomics, type 2 diabetes, and obesity. *The New England journal of medicine* **363**, 2339-50 (2010).

130. Lawless, M., et al.Targeting epigenetic pathways in obesity and diabetes. **4**, (2008).

131. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396-398 (2007).

132. Cubas P., et al. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157-161 (1999).

133. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693-705 (2007).

134. Haberland M., et al. The many roles of histone deacetylases in development and physiology: Implications for disease and therapy. *Nature Reviews Genetics* **10**, 32-42 (2009).

135. Wang Z., et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics* **40**, 897-903 (2008).

136. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-37 (2007).

137. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).

138. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).

139. Roh, T.-Y., et al. The genomic landscape of histone modifications in human T cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15782-7 (2006).

140. Bhandare, R. et al. Genome-wide analysis of histone modifications in human pancreatic islets. *Genome research* **20**, 428-33 (2010).

141. Hermann, A., et al. Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol. Life Sci.* **61**, 2571-87 (2004).

142. Okano, M., et al. Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57 (1999).

143. Clouaire T., et al. Methyl-CpG binding proteins: Specialized transcriptional repressors or structural components of chromatin? *Cellular and Molecular Life Sciences* **65**, 1509-1522 (2008).

144. Takahashi T., et al. Aberrant promoter methylation of multiple genes during multistep pathogenesis of colorectal cancers. *International Journal of Cancer* **118**, 924-931 (2006).

145. Feltus F.A., et al. Predicting aberrant CpG island methylation. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12253-12258 (2003).

146. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature reviews. Genetics* **10**, 295-304 (2009).

147. Schübeler D., et al. Genomic targeting of methylated DNA: Influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Molecular and Cellular Biology* **20**, 9103-9112 (2000).

148. En Li. Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Reviews Genetics* **3**, 662-73 (2002).

149. Wolff G.L., et al. Maternal epigenetics and methyl supplements affect agouti gene expression in A(vy)/a mice. *FASEB Journal* **12**, 949-957 (1998).

150. Cooney C.A., et al. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *Journal of Nutrition* **132**, 2393S-2400S (2002).

151. de Rooij, SR, et al. Impaired insulin secretion after prenatal exposure to the Dutch famine. *Diabetes Care* **29**, 1897-901 (2006).

152. Gluckman, P.D. et al. Metabolic plasticity during mammalian development is directionally dependent on early nutritional status. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12796-800 (2007).

153. Simmons R.A., et al. Progressive accumulation of mitochondrial DNA mutations and decline in mitochondrial function lead to β-cell failure. *Journal of Biological Chemistry* **280**, 28785-28791 (2005).

154. Simmons R.A., et al. Intrauterine Growth Retardation Leads to the Development of Type 2 Diabetes in the Rat. *Diabetes* **50**, 2279-2286 (2001).

155. Betz, R., et al. Human histone deacetylase 2, HDAC2 (Human RPD3), is localized to 6q21 by radiation hybrid mapping. *Genomics* **42**, 245-6 (1998).

156. Gray S.G., E.T.J. The human histone deacetylase family. *Experimental Cell Research* **262**, 75-83 (2001).

157. Yamauchi T., et al. Increased insulin sensitivity despite lipodystrophy in Crebbp heterozygous mice. *Nature Genetics* **30**, 221-226 (2002).

158. Johnstone, S.E. & Baylin, S.B. Stress and the epigenetic landscape: a link to the pathobiology of human diseases? *Nature reviews. Genetics* **11**, 806-12 (2010).

159. Drake J., et al. 4-Hydroxynonenal oxidatively modifies histones: Implications for Alzheimer's disease. *Neuroscience Letters* **356**, 155-158 (2004).

160. Hitchler, M.J., Domann, F.E. An epigenetic perspective on the free radical theory of development. *Free Radic. Biol. Med* **43**, 1023-36 (2007).

161. Mosley, A.L., et al. The pancreatic duodenal homeobox-1 protein (Pdx-1) interacts with histone deacetylases Hdac-1 and Hdac-2 on low levels of glucose. *Journal of Biological Chemistry* **279**, 54241-54247 (2004).

162. Deering, T.G., et al. Methyltransferase Set7/9 maintains transcription and euchromatin structure at islet-enriched genes. *Diabetes* **58**, 185-93 (2009).

163. Mosley, A.L. & Ozcan, S. Glucose regulates insulin gene transcription by hyperacetylation of histone h4. *The Journal of biological chemistry* **278**, 19660-6 (2003).

164. Wang, H.-W., et al. Pdx-1 modulates histone H4 acetylation and insulin gene expression in terminally differentiated alpha-TC-1 cells. *Pancreas* **34**, 248-53 (2007).

165. Gerrish K., et al. Conserved transcriptional regulatory domains of the pdx-1 gene. *Molecular Endocrinology* **18**, 533-548 (2004).

166. Qiu Y.I., et al. p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. *Molecular and Cellular Biology* **18**, 2957-2964 (1998).

167. Park J.H., et al. Development of type 2 diabetes following intrauterine growth retardation in rats is associated with progressive epigenetic silencing of Pdx1. *Journal of Clinical Investigation* **118**, 2316-2324 (2008).

168. Kaiser N., et al. Psammomys obesus, a model for environment-gene interactions in type 2 diabetes. *Diabetes* **54**, S137-S144 (2005).

169. Poitout, V. & Robertson, R.P. Glucolipotoxicity: fuel excess and beta-cell dysfunction. *Endocrine reviews* **29**, 351-66 (2008).

170. Liu, W.-D., et al. INSM1 functions as a transcriptional repressor of the neuroD/beta2 gene through the recruitment of cyclin D1 and histone deacetylases. *The Biochemical journal* **397**, 169-77 (2006).

171. Seol W., et al. An orphan nuclear hormone receptor that lacks a DNA binding domain and heterodimerizes with other receptors. *Science* **272**, 1336-1339 (1996).

172. Lee Y.-K., et al. The orphan nuclear receptor SHP inhibits hepatocyte nuclear factor 4 and retinoid X receptor transactivation: Two mechanisms for repression. *Molecular and Cellular Biology* **20**, 187-195 (2000).

173. Kim J.-Y., et al. Orphan nuclear receptor small heterodimer partner, a novel corepressor for a basic helix-loop-helix transcription factor BETA2/NeuroD. *Molecular Endocrinology* **18**, 776-790 (2004).

174. Sproul D., et al. The role of chromatin structure in regulating the expression of clustered genes. *Nature Reviews Genetics* **6**, 775-781 (2005).

175. Palstra, R., et al. The b-globin nuclear compartment in development and erythroid differentiation. *Nature genetics* **35**, 190-194 (2003).

176. Mutskov, V. & Felsenfeld, G. The human insulin gene is part of a large open chromatin domain specific for human islets. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 17419-24 (2009).

177. Gaulton, K.J. et al. A map of open chromatin in human pancreatic islets. *Nature genetics* **42**, 255-9 (2010).

178. Grant S. et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nature Genetics* **38**, 320-323 (2006).

179. Stitzel, M.L. et al. Global Epigenomic Analysis of Primary Human Pancreatic Islets Provides Insights into Type 2 Diabetes Susceptibility Loci. *Cell Metabolism* 443-455 (2010).

180. Lynn, F. Meta-regulation: microRNA regulation of glucose and lipid metabolism. *Trends Endocrinol. Metab.* **20**, 452-9 (2009).

181. Groop, L. Open chromatin and diabetes risk. *Nature genetics* **42**, 190-2 (2010).

182. Robertson R.P., et al. Glucose toxicity in β-cells: Type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes* **52**, 581-587 (2003).

183. Unger, R.H. Minireview: Weapons of Lean Body Mass Destruction: The Role of Ectopic Lipids in the Metabolic Syndrome. *Endocrinology* **144**, 5159-5165 (2003).

184. Poitout V. Minireview: Secondary β-cell failure in type 2 diabetes - A convergence of glucotoxicity and lipotoxicity. *Endocrinology* **143**, 339-342 (2002).

185. Milne, J.C. & Denu, J.M. The Sirtuin family: therapeutic targets to treat diseases of aging. *Current opinion in chemical biology* **12**, 11-7 (2008).

186. Cho, Y.M. et al. Betacellulin and nicotinamide sustain PDX1 expression and induce pancreatic beta-cell differentiation in human embryonic stem cells. *Biochemical and biophysical research communications* **366**, 129-34 (2008).

187. Ingenuity Systems Ingenuity Pathway Analysis. http://www.ingenuity.com

188. Giresi, P.G., et al. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research* **17**, 877-85 (2007).

189. Heinz, S., et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* **38**, 576-89 (2010).

190. Boyle, A.P., et al. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* **24**, 2537-2538 (2008).

191. Aramata S., et al. Synergistic activation of the insulin gene promoter by the β-cell enriched transcription factors MafA, Beta2, and Pdx1. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1730**, 41-46 (2005).

192. Ohneda K., et al. Regulation of insulin gene transcription. *Seminars in Cell and Developmental Biology* **11**, 227-233 (2000).

193. Melloul, D., et al. Regulation of insulin gene transcription. *Diabetologia* **45**, 309-26 (2002).

194. Brissova, M. et al. Reduced PDX-1 expression impairs islet response to insulin resistance and worsens glucose homeostasis. *American journal of physiology. Endocrinology and metabolism* **288**, E707-14 (2005).

195. Stoffers D.A., et al. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nature genetics* **17**, 138-139 (1997).

196. Al-Quobaili, F. & Montenarh, M. Pancreatic duodenal homeobox factor-1 and diabetes mellitus type 2. *International Journal of Molecular Medicine* **29**, 399-404 (2008).

197. Martin, M., et al. Transcription factors in pancreatic development. Animal models. *Endocrine development* **12**, 24-32 (2007).

198. Brink, C. Promoter elements in endocrine pancreas development and hormone regulation. *Cellular and molecular life sciences : CMLS* **60**, 1033-48 (2003).

199. Sharma A., et al. The role of the insulin control element and RIPE3b1 activators in glucose- stimulated transcription of the insulin gene. *Molecular Endocrinology* **9**, 1468-1476 (1995).

200. Itkin-Ansari, P., et al. Cell-based therapies for diabetes: progress towards a transplantable human beta cell line. *Annals of the New York Academy of Sciences* **1005**, 138-147 (2003).

201. Tour, D.D. et al. β-Cell Differentiation from a Human Pancreatic Cell Line in Vitro and in Vivo. *Molecular Endocrinology* **15**, 476 (2001).

02. Narushimaya, M., et al. A human beta-cell line for transplantation therapy to control type 1 diabetes. *Nature biotechnology* **10**, 1274-82 (2005).

203. Rivas-Carrill, J.D., et al. Current Cell-based Approaches for the Treatment of Diabetes Mellitus. *Biotechnology and Genetic Engineering Reviews* **24**, 287-96 (2007).

204. LeRoith, D., Olefsky, J.M. & Taylor, S.I. *Diabetes Mellitus: A Fundamental and Clinical Text*. (2003).

205. Asfari, M., Janjic, D. & Meda, P. Establishment of 2-mercaptoethanol-dependent differentiated insulin- secreting cell lines. *Endocrinology* **130**, 167-68 (1992).

206. Miyazaki, J., Araki, K. & Yamato, E. Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* **127**, 126-132 (1990).

207. Minami, K. et al. Insulin secretion and differential gene expression in glucose-responsive and -unresponsive MIN6 sublines. *American journal of physiology. Endocrinology and metabolism* **279**, E773-81 (2000).

208. Hamaguchi, K., Gaskins, H.R. & Leiter, E.H. NIT-1, a pancreatic beta-cell line established from a transgenic NOD/Lt mouse. *Diabetes* **7**, 842-49 (1991).

209. Struhl, K. Interpreting chromatin immunoprecipitation results. *Cell* 29-34 (2008).

210. Bergholdt, R., et al. Characterization of new polymorphisms in the 5' UTR of the human interleukin-1 receptor type 1 (IL1R1) gene: linkage to type 1 diabetes and correlation to IL-1RI plasma level. *Genes and Immunity* **1**, 495-500 (2000).

211. Robinson, J.T. et al. Intergrative Genomics Viewer. *Nature Biotechnology* **29**, 24-26 (2011).

212. Thomas, P.D. et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research* **34**, W645-W650 (2006).

213. Park, P.J. ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics* **10**, 669-80 (2009).

214. Pan, N., et al. Defects in the cerebella of conditional Neurod1 null mice correlate with effective Tg(Atoh1-cre) recombination and granule cell requirements for Neurod1 for differentiation. *Cell and Tissue Research* **337**, 407-28 (2009).

215. Klischa, T.J. et al. In vivo Atoh1 targetome reveals how a proneural transcription factor regulates cerebellar development. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3288-93 (2011).

216. Aronheim, A., Shiran, R., Rosen, A. & Walker, M.D. The E2A gene product contains two separable and functionally distinct transcription activation domains. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8063-7 (1993).

217. Cordle, S.R., Henderson, E., Masuoka, H., Weil, P.A. & Stein, R. Pancreatic beta-cell-type-specific transcription of the insulin gene is mediated by basic helix-loop-helix DNA-binding proteins. *Molecular and cellular biology* **11**, 1734-38 (1991).

218. Firulli, A.B. & Olson, E.N. Modular regulation of muscle gene transcription: a mechanism for muscle cell diversity. *Trends in Genetics* **13**, 364-69 (1997).

219. Sommer, L., Ma, Q. & Anderson., D.J. Neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. *Molecular and Cellular Neuroscience* **8**, 221-41 (1996).

220. Grindstaff, K.K., et al. Sec6/8 complex is recruited to cell-cell contacts and specifies transport vesicle delivery to the basal-lateral membrane in epithelial cells. *Cell* **93**, 731-40 (1998).

221. Laramie, J.M., et al. Polymorphisms near EXOC4 and LRGUK on chromosome 7q32 are associated with Type 2 Diabetes and fasting glucose; the NHLBI Family Heart Study. *BMC medical genetics* **9**, 46 (2008).

222. Brancaccio, M., et al. Melusin is a new muscle-specific interactor for beta(1) integrin cytoplasmic domain. *Journal of Biological Chemistry* **274**, 29282-8 (1999).

223. Twu, Y.C., et al. I branching formation in erythroid differentiation is regulated by transcription factor C/EBPalpha. *Blood* **110**, 4526-34 (2007).

224. Chen, B.-Z. et al. Identification of microRNAs expressed highly in pancreatic islet-like cell clusters differentiated from human embryonic stem cells. *Cell Biology International* **31**, 29-37 (2011).

225. Muraoka, M. et al. Involvement of SIK2/TORC2 signaling cascade in the regulation of insulin-induced PGC-1alpha and UCP-1 gene expression in brown adipocytes. *American Journal of Physiology - Endocrinology and Metabolism* **296**, 1430-9 (2006).

226. Delibegovic, M. et al. Disruption of the Striated Muscle Glycogen Targeting Subunit PPP1R3A of Protein Phosphatase 1 Leads to Increased Weight Gain, Fat Deposition, and Development of Insulin Resistance. *Diabetes* **52**, 596-604 (2003).

227. Xia, J., et al. A common variant in PPP1R3 associated with insulin resistance and type 2 diabetes. *Diabetes* **47**, 1519-24 (1998).

228. Ishida, S. et al Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol Cell Biochem* **21**, 4684–469 (2001).

229. Kremer, B., et al. Septins regulate actin organization and cell-cycle arrest through nuclear accumulation of NCK mediated by SOCS7. *Cell* **130**, 837-50 (2007).

230. Zhu, M., et al. Septin 7 interacts with centromere-associated protein E and is required for its kinetochore localization. *Journal of Biological Chemistry* **283**, 18916-25 (2008).

231. Bouteldja, M., et al. The biochemical basis of hereditary fructose intolerance. *Journal of Inherited Metabolic Diseases* **33**, 105-12 (2010).

232. Wang H., et al. Hepatocyte Nuclear Factor 4α Regulates the Expression of Pancreatic β-Cell Genes Implicated in Glucose Metabolism and Nutrient-induced Insulin Secretion. *Journal of Biological Chemistry* **275**, 35953-9 (2000).

233. Marzluff, M., et al. The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-98 (2002).

234. YuA, U. et al. Human Na+,K+-ATPase genes. Beta-subunit gene family contains at least one gene and one pseudogene. *FEBS Letters* **257**, 439-42 (1999).

235. Melzer, N., et al. Glutamate modifies ion conduction and voltage-dependent gating of excitatory amino acid transporter-associated anion channels. *Journal of Biological Chemistry* **278**, 50112-9 (2003).

236. Cairano, D.C., et al. The glial glutamate transporter 1 (GLT1) is expressed by pancreatic beta-cells and prevents glutamate-induced beta-cell death. *Journal of Biological Chemistry* **286**, 14007-18 (2011).

237. Cai, H. et al. Complexin II plays a positive role in Ca2+-triggered exocytosis by facilitating vesicle priming. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19538-43 (2008).

238. Smith, K.A., et al. Interactions between factor XIII and the alphaC region of fibrinogen. *Blood* **117**, 3460-8 (2011).

239. Nakamura, T., et al. N-Shc and Sck, two neuronally expressed Shc adapter homologs. Their differential regional expression in the brain and roles in neurotrophin and Src signaling. *Journal of Biological Chemistry* **273**, 6960-7 (1998).

240. Watada H.,et al. Transcriptional and translational regulation of β-cell differentiation factor Nkx6.1. *Journal of Biological Chemistry* **275**, 34224-34230 (2000).

241. Odom, D.T. et al. Control of pancreas and liver gene expression by HNF transcription factors. *Science (New York, N.Y.)* **303**, 1378-81 (2004).

242. Lee Y.-H., et al. Laron dwarfism and non-insulin-dependent diabetes mellitus in the Hnf- 1α knockout mouse. *Molecular and Cellular Biology* **18**, 3059-3068 (1998).

243. Wang H., et al. Dominant-negative suppression of HNF-1α function results in defective insulin gene transcription and impaired metabolism-secretion coupling in a pancreatic β-cell line. *EMBO Journal* **17**, 6701-6713 (1998).

244. Gunton J.E., et al. Loss of ARNT/HIF1β mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* **122**, 337-349 (2005).

245. Narita R., et al. Islet amyloid polypeptide (IAPP) and pancreatic islet amyloid depostition in diabetic and non-diabetic patients. *Diabetes Research and Clinical Practice* **15**, 3-14 (1992).

246. Cui, L., et al. Cloning and characterization of islet cell antigen-related protein-tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen in insulin-dependent diabetes. *Journal of Biological Chemistry* **271**, 24817-23 (1994).

247. Solimena, M., et al. ICA 512, an autoantigen of type I diabetes, is an intrinsic membrane protein of neurosecretory granules. *EMBO Journal* **15**, 2102-4 (1994).

248. Shen, L., et al. NDRG2 is highly expressed in pancreatic beta cells and involved in protection against lipotoxicity. *Cell Mol. Life Sci.* **67**, 1371-81 (2010).

249. Skelly, R., et al. Glucose-regulated translational control of proinsulin biosynthesis with that of the proinsulin endopeptidases PC2 and PC3 in the insulin-producing MIN6 cell line. *Diabetes* **45**, 37-43 (1996).

250. Sun, C., et al. Kirrel2, a novel immunoglobulin superfamily gene expressed primarily in beta cells of the pancreatic islets. *Genomics* **82**, 130-42 (2003).

251. Wang, J. et al. A mutation in the insulin 2 gene induces diabetes with severe pancreatic β-cell dysfunction in the Mody mouse. *Journal of Clinical Investigation* **103**, 27-37 (1999).

252. Guillam, M.T. et al. Early diabetes and abnormal postnatal pancreatic islet development in mice lacking Glut-2. *Nature Genetics* **17**, 327-30 (1997).

253. Akpinar, P., et al. Tmem27: a cleaved and shed plasma membrane protein that stimulates pancreatic beta cell proliferation. *Cell Metabolism* **6**, 385-97 (2005).

254. Mayr, A., et al. GAD autoantibody affinity and epitope specificity identify distinct immunization profiles in children at risk for type 1 diabetes. *Diabetes* **56**, 1527-33 (2007).

255. Lynedjian, P.B. Molecular physiology of mammalian glucokinase. *Cell Mol. Life Sci.* **66**, 27-42 (2009).

256. Rizo, J., et al. Snares and Munc18 in synaptic vesicle fusion. *Nature Reviews Neuroscience* **8**, 641-53 (2002).

257. Bennet, M.K., et al. Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. *Science* **257**, 255-9 (1992).

258. Striggow, F. & Ehrlich, B.E. Ligand-gated calcium channels inside and out. *Current Opinion in Cell Biology* **8**, 490-5 (1996).

259. Yi, Z., et al. The Rab27a/granuphilin complex regulates the exocytosis of insulin-containing dense-core granules. *Molecular and Cellular Biology* **22**, 1858-67 (2002).

260. Gloyn, A.L., et al. Mutations in the genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Human Mutation* **27**, 220-31 (2006).

261. Gustavsson, N. et al. Impaired insulin secretion and glucose intolerance in synaptotagmin-7 null mutant mice. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3992-7 (2008).

262. Brownlee, M. A radical explanation for glucose-induced beta cell dysfunction. *The Journal of clinical investigation* **112**, 1788-90 (2003).

263. Xu, G. et al. Insulin Regulation of β-Cell Function Involves a Feedback Loop on SERCA Gene Expression, Ca2+ Homeostasis, and Insulin Expression and Secretion. *Biochemistry* **39**, 14912–14919 (2000).

264. Obermüller, S. et al. Defective secretion of islet hormones in chromogranin-B deficient mice. *PLoS ONE* **28**, e8936 (2010).

265. Curry, D.L., et al. Requirement for calcium ion in insulin secretion by the perfused rat pancreas. *AM.J.PHYSIOL.* **214**, 174-8 (1968).

266. Graves, T.K., et al. Ca2+-induced Ca2+ release in the pancreatic β-cell: direct evidence of endoplasmic reticulum Ca2+ release. *Endocrinology* **144**, 3565-74 (2003).

267. Pabst, S. et al. Rapid and selective binding to the synaptic SNARE complex suggests a modulatory role of complexins in neuroexocytosis. *Journal of Biological Chemistry* **277**, 7838-48 (2001).

268. Ono, S., et al. Regulatory roles of complexins in neurotransmitter release from mature presynaptic nerve terminals. *European Journal of Neuroscience* **10**, 2143-52 (1998).

269. Itakura, M., et al. Transfection analysis of functional roles of complexin I and II in the exocytosis of two different types of secretory vesicles. *Biochemical and Biophysical Research Communications* **265**, 691-96 (1999).

270. Huntwork, S., et al. A complexin fusion clamp regulates spontaneous neurotransmitter release and synaptic growth. *Nature Neuroscience* **10**, 1235-7 (2007).

271. Abderrahmani, A. et al. Complexin I regulates glucose-induced secretion in pancreatic beta-cells. *Journal of cell science* **117**, 2239-47 (2004).

272. MacMahon, T., et al. Complexins: cytosolic proteins that regulate SNAP receptor function. *Cell* **83**, 111-9 (1995).

273. Eastwood, S.L., et al. Expression of complexin I and II mRNAs and their regulation by antipsychotic drugs in the rat forebrain. SYNAPSE. *Synapse* **36**, 166-7 (2000).

274. Iglesias, A. et al. Diabetes and exocrine pancreatic insufficiency in E2F1 / E2F2 double-mutant mice. *Span* **113**, (2004).

275. Kimple, M.E. et al. Gαz Negatively Regulates Insulin Secretion and Glucose Clearance. *Journal of Biological Chemistry* **283**, 4560-7 (2007).

276. Leahy J.L., et al. β-Cell dysfunction induced by chronic hyperglycemia: Current ideas on mechanism of impaired glucose-induced insulin secretion. *Diabetes Care* **15**, 442-455 (1992).

277. Biden, T.J., et al. Chronic effects of fatty acids on pancreatic beta-cell function: new insights from functional genomics. *Diabetes* **53**, S159–S165 (2004).

278. Glauser, D. et al. Transcriptional response of pancreatic beta cells to metabolic stimulation: large scale identification of immediate-early and secondary response genes. *BMC molecular biology* **8**, 54 (2007).

279. Kelpe, C.L., et al. Palmitate inhibition of insulin gene expression is mediated at the transcriptional level via ceramide synthesis. *Journal of Biological Chemistry* **278**, 30015–21 (2003).

280. Hagman D.K.,et al. Palmitate inhibits insulin gene expression by altering PDX-1 nuclear localization and reducing MafA expression in isolated rat islets of Langerhans. *Journal of Biological Chemistry* **280**, 32413-32418 (2005).

281. Olofsson, C.S. et al. Long-Term Exposure to Glucose and Lipids Inhibits Glucose-Induced Insulin Secretion Downstream of Granule Fusion With Plasma Membrane. **56**, (2007).

282. Schuit, F., et al. Glucose-Regulated Gene Expression Maintaining the Glucose-Responsive State of -Cells. *Diabetes* **51**, 326S-332 (2002).

283. Brutlag, D., Schlehuber, C. & Bonner, J. Properties of formaldehyde-treated nucleohistone. *Biochemistry* **8**, 3214–3218 (1969).

284. Polach, K.J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *Journal of molecular biology* **254**, 130–149 (1995).

285. Yamagata, K., et al. Voltage-gated K+ channel KCNQ1 regulates insulin secretion in MIN6 β-cell line. *Biochemical and Biophysical Research Communications* **407**, 620-5 (2011).

286. Xavier, G., et al. TCF7L2 Regulates Late Events in Insulin Secretion From Pancreatic Islet Beta-Cells. *Diabetes* **58**, 894-905 (2009).

287. Artner, I., et al. MafB is an activator of glucagon gene expression in alpha- and beta cells. *Diabetes* **55**, 297-304 (2006).

288. Sakurai, T., et al. Orexins and orexin receptors: a family of hypothalamic neuropeptides and G-protein coupled receptors that regulate feeding behaviour. *Cell* **92**, (1998).

289. Ouedraogo, R., Näslund, E. & Kirchgessner, A.L. Glucose Regulates the Release of Orexin-A From the Endocrine Pancreas. *Diabetes* **52**, 111-7 (2003).

290. Adeghate, E. et al. Orexin-1 Receptor Co-Localizes with Pancreatic Hormones in Islet Cells and Modulates the Outcome of Streptozotocin-Induced Diabetes Mellitus. *PLoS ONE* **5**, e8587 (2010).

291. Vanderford, N.L., et al. Multiple Kinases Regulate mafA Expression in the Pancreatic Beta Cell Line MIN6. *Archives of biochemistry and biophysics* **480**, 138-142 (2008).

292. Plaisance, V. et al. MicroRNA-9 controls the expression of Granuphilin/Slp4 and the secretory response of insulin-producing cells. *The Journal of biological chemistry* **281**, 26932-42 (2006).

293. Ramachandran, D. et al. Sirt1 and mir-9 expression is regulated during glucose-stimulated insulin secretion in pancreatic β-islets. *FEBS letters* **278**, 1167-74 (2011).

294. Bordone, L. et al. Sirt1 regulates insulin secretion by repressing UCP2 in pancreatic beta cells. *PLoS biology* **4**, e31 (2006).

295. Hudson, C., et al. Sox17α and -β Mediate Endoderm Formation in Xenopus. *Cell* **91**, 397-405 (1997).

296. Sinner, D., et al. Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development* **131**, 3069-80 (2004).

297. Lioubinski, O., Muller, M., Wegner, M. & Sander, M. Expression of Sox Transcription Factors in the Developing Mouse Pancrea. *Developmental Dynamics* **227**, 402-8 (2003).

298. Wells, J. The function of Sox 17 in the adult pancreas. at <http://www.cincinnatichildrens.org/research/div/dev-biology/fac-labs/wells-lab/adult-stem-cells.htm>

299. Kim, H., et al. Epigenetic signatures and temporal expression of lineage-specific genes in hESCs during differentiation to hepatocytes in vitro. *Human Molecular Genetics* **20**, 401-12 (2011).

300. Takahashi, K., et al. Expression of the semaphorins Sema 3D and Sema 3F in the developing parathyroid and thymus. *Developmental Dynamics* **237**, 1699-708 (2008).

301. Wu, Q. & Maniatis, T. A striking organization of a large family of human neural cadherinlike cell adhesion genes. *Cell* **97**, 779-90 (1999).

302. Sherwood, R.I. et al. Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Developmental biology* **304**, 541-55 (2007).

303. Kaneda, H., et al. FOXQ1 is overexpressed in colorectal cancer and enhances tumorigenicity and tumor growth. *Cancer Research* **70**, 2053-63 (2010).

304. Low, S.-K. et al. Genome-Wide Association Study of Pancreatic Cancer in Japanese Population. *PloS one* **5**, e11824 (2010).

305. Aitola, M., Carlsson, P., Mahlapuu, M., Enerbäck, S. & Pelto-Huikko, M. Forkhead transcription factor FoxF2 is expressed in mesodermal tissues involved in epithelio-mesenchymal interactions. *Developmental Dynamics* **218**, 136-149 (2000).

306. Perney, T.M. & Kaczmarek, L.K. Expression and regulation of mammalian K+ channel genes. *Seminars in Neuroscience* **5**, 135-145 (1999).

307. Jesperson, T., et al. The KCNQ1 potassium channel: from gene to physiological function. *Physiology* **20**, 408-16 (2005).

308. Yasuda, K. et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nature Genetics* **40**, 1092-7 (2008).

309. A, J. et al. A variant in the KCNQ1 gene predicts future type 2 diabetes and mediates impaired insulin secretion. *Diabetes* **58**, 2049-13 (2008).

310. Mancini-DiNardo, D., et al. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes and Development* **20**, 1268-82 (2006).

311. Pandey, R.R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular Cell* **32**, 232-42 (2008).

312. Miao, F. et al. Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes. *Diabetes* **57**, 3189-98 (2008).

313. Nicolas, M., et al. Transcriptional regulation of the human Sp1 gene promoter by the specificity protein (Sp) family members nuclear factor Y (NF-Y) and E2F. *Biochemical Journal* **371**, 265-75 (2003).

314. Solomon, S.S., et al. A critical role of Sp1 transcription factor in regulating gene expression in response to insulin and other hormones. *Life Science* **83**, 305-12 (2008).

315. Dunn, K.L., et al. The many roles of the transcriptional regulator CTCF. *Biochemistry and cellular biology* **81**, 161-7 (2003).

316. Mantovani, R. The molecular biology of the CCAAT-binding factor NF-Y. *Gene* **239**, 15-27 (1999).

317. Vatamaniuk, M.Z. et al. Foxa1-Deficient Mice Exhibit Impaired Insulin Secretion due to Uncoupled Oxidative Phosphorylation. *Diabetes* **55**, 2730-6 (2006).

318. Gao, N. et al. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes & development* **22**, 3435-48 (2008).

319. Ojeda, S.R. et al. The Oct-2 POU Domain Gene in the Neuroendocrine Brain: A Transcriptional Regulator of Mammalian Puberty. *Endocrinology* **140**, 3774 (1999).

320. Loh, Y.-H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics* **38**, 431-440 (2006).

321. Wen, J., et al. Oct4 and Nanog expression is associated with early stages of pancreatic carcinogenesis. *Pancreas* **39**, 622-6 (2010).

322. Bottino, R. et al. Response of Human Islets to Isolation Stress and the Effect of Antioxidant Treatment. *Diabetes* **53**, (2004).

323. Wang, H., et al. ER stress and SREBP-1 activation are implicated in beta -cell glucolipotoxicity. *Journal of cell science* **118**, 3905–15 (2005).

324. Chang-Chen, K.J., Mullur, R. & Bernal-Mizrachi, E. Beta-cell failure as a complication of diabetes. *Reviews in endocrine & metabolic disorders* **9**, 329-43 (2008).

325. Hoffman, B.G. et al. Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. *Genome research* **20**, 1037-51 (2010).

326. Rodriguez, J. et al. Bivalent Domains Enforce Transcriptional Memory of DNA Methylated Genes in Cancer Cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19809-19814 (2008).

327. Murry, C.E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661-80 (2008).

328. Soria, B. et al. Insulin-Secreting Cells Derived From Embryonic Stem Cells Normalize Glycemia in Streptozotocin-Induced Diabetic Mice. *Diabetes* **49**, 157-162 (2000).

329. Soria, B. In-vitro differentiation of pancreatic beta-cells. *Differentiation; research in biological diversity* **68**, 205-19 (2001).

330. Lumelsky N., et al. Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science* **292**, 1389-1394 (2001).

331. Hori Y., et al. Growth inhibitors promote differentiation of insulin-producing tissue from embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16105-16110 (2002).

332. Blyszczuk P., et al. Expression of Pax4 in embryonic stem cells promotes differentiation of nestin-positive progenitor and insulin-producing cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 998-1003 (2003).

333. Kyba, M., Perlingeiro, R.C.R. & Daley, G.Q. HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* **109**, 29-37 (2002).

334. Serafimidis, I., Rakatzi, I., Episkopou, V., Gouti, M. & Gavalas, A. Novel effectors of directed and Ngn3-mediated differentiation of mouse embryonic stem cells into endocrine pancreas progenitors. *Stem cells (Dayton, Ohio)* **26**, 3-16 (2008).

335. Hori, Y., Gu, X., Xie, X. & Kim, S.K. Differentiation of insulin-producing cells from human neural progenitor cells. *PLoS medicine* **2**, e103 (2005).

336. Yasunaga, M. et al. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nature biotechnology* **23**, 1542-50 (2005).

337. Nikolova, G. et al. The vascular basement membrane: a niche for insulin gene expression and Beta cell proliferation. *Developmental cell* **10**, 397-405 (2006).

338. Jiang, F.X., Cram, D.S., DeAizpurua, H.J. & Harrison, L.C. Laminin-1 promotes differentiation of fetal mouse pancreatic beta-cells. *Diabetes* **48**, 722-30 (1999).

339. Weber, L.M., Hayda, K.N. & Anseth, K.S. Cell-matrix interactions improve beta-cell survival and insulin secretion in three-dimensional culture. *Tissue engineering. Part A* **14**, 1959-68 (2008).

340. Kleinman, H. K., et al. Review Matrigel: Basement membrane matrix with biological activity. *Seminars in cancer biology* **15**, 378-86 (2005).

341. Oberg-Welsh, C. Long-term culture in matrigel enhances the insulin secretion of fetal porcine islet-like cell clusters in vitro. *Pancreas* **22**, 157-63 (2001).

342. Johansson, M., Mattsson, G., Andersson, A., Jansson, L. & Carlsson, P.-O. Islet endothelial cells and pancreatic beta-cell proliferation: studies in vitro and during pregnancy in adult rats. *Endocrinology* **147**, 2315-24 (2006).

343. Movassat, J. Exendin 4 Up-Regulates Expression of PDX 1 and Hastens Differentiation and Maturation of Human Fetal Pancreatic Cells. *Journal of Clinical Endocrinology & Metabolism* **87**, 4775-4781 (2002).

344. Tayaramma, T., Ma, B., Rohde, M. & Mayer, H. Chromatin-remodeling factors allow differentiation of bone marrow cells into insulin-producing cells. *Stem cells (Dayton, Ohio)* **24**, 2858-67 (2006).

345. Haumaitre, C., Lenoir, O., Scharfmann, R. & Haumaitre C., L.O.S.R. Histone deacetylase inhibitors modify pancreatic cell fate determination and amplify endocrine progenitors. *Molecular and Cellular Biology* **28**, 6373-6383 (2008).

346. Aramata S., et al. Roles and regulation of transcription factor MafA in islet beta-cells. *Endocrine journal* **54**, 659-66 (2007).

347. Newsholme, P., et al. Nutrient regulation of insulin secretion and beta-cell functional integrity. *Advances in Experimental Medicine and Biology* **654**, 91-114 (2010).

348. Santini, V., Kantarjian, H. & Issa, J. Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. *Ann Intern Med.* **134**, 573-86 (2001).

349. Juttermann, R., Li, E. & Jaenish, R. Toxicity of 5-aza-2'deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 11797-801 (1994).

350. Cameron, E.E., Bachman, K.E., Myohanen, S., J.G., H. & Baylin, S.B. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nature Genetics* **21**, 103-7 (1999).

351. Gerhart-Hines, J., et al. Metabolic control of muscle mitochondrial function and fatty acid oxidation through SIRT1/PGC-1alpha. *EMBO Journal* **26**, 1913-23 (2007).

352. Yoshizaki, T., et al. SIRT1 inhibits inflammatory pathways in macrophages and modulates insulin sensitivity. *Am J Physiol Endocrinol Metab* **298**, E419-428 (2010).

353. Yoshizaki, T., et al. SIRT1 exerts anti-inflammatory effects and improves insulin sensitivity in adipocytes. *Molecular and cellular biology* **29**, 1363-1374 (2009).

354. Sirtris Pharmaceuticals. at <http://www.sirtrispharma.com/discovery-sirt1.html>

355. Bordone, L., et al. SIRT1 transgenic mice show phenotypes resembling calorie restriction. *Aging Cell* **6**, 759-67 (2007).

356. Qi, Y. et al. Sirtris Pharmaceuticals: Abstracts to be presented at the 71st Scientific Sessions of the American Diabetes Association Meeting in San Diego, CA, in June 2011. at <http://www.sirtrispharma.com/scientific-meetings.html>